Privacy & Monetization on the Web

Abstract:

The impact of a decentralized web and the proliferation of access methods will redefine how services and resources are subsidized by advertising. The impact of AI at the edge will move smart computing towards, devices, smartphones and non-traditional access points to the Internet. Privacy will impact how existing data aggregators share information such as advertisers. Users will enter into the equation to monetize their private data so advertisers and other parties can reward them for engagement.

UbiVault

Background

Internet use for commercial applications is rapidly expanding due to innovation as well as evolution of technology.

In the past, advertisers sought to deliver impressions within a website, nominally as part of the viewed content. Advertisers had to pay the web publisher to be visible, using data aggregators such as Google and others to provide better targeting for search engine optimization (SEO) as well as content relevancy; in other words, placing ads where appropriate.

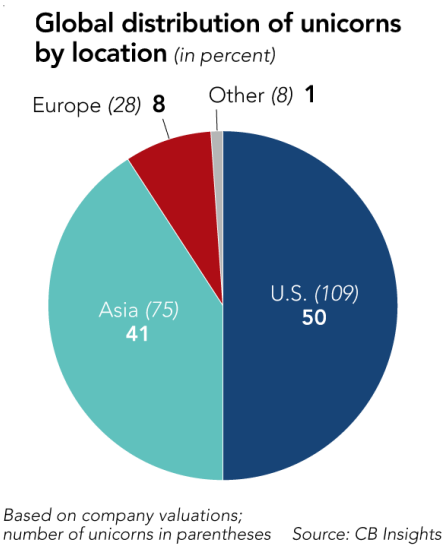
This has led to an entire industry of collecting data on web activity provided by data aggregators to improve targeted response. Knowing something about the viewer is the key requirement. Its all about the "who, what, when, where and why (propensity to react)".

On the other side of the debate is the issue of privacy which today, on the Web is a chimera. Many vendors provide tools that leverage the knowledge of where you physically are when you engage (GPS), what you interacted with at any point in time (profiles) and viewing metrics that price your activity (propensity to respond) - all in real time. It's the principle behind delivery of advertising through digital media networks that allow automatic bidding of ad placement.

The problem is that the Internet is changing from a server centric model to a peer-to-peer (P2P) network. New methods allow people to create virtual networks of relationships that do not require a centralized application, typically on a cloud resource.

The current world of Facebook, Google+ (now defunct), LinkedIn and others, is being replaced by virtual relationships that are easy to implement. This shows the existing web publishing model is about to change. In simple terms, anyone can be their own virtual website of interests. Advertisers need to adapt.

The growing interest by investors in the new Internet (web 3.0), dominated by the new unicorns in Asia, shows momentum shifting in how advertisers reach consumers. Asia now has 75 unicorn startups, accounting for 40% of the global total by value, a feat made possible by the region's booming on-demand industry.



China's Didi has 400 million users of its ride-hailing app. Smartphone Xiaomi, drone maker DJI, bike-sharing giant Mobike, and China's answer to Airbnb -- Tujia. All are now developing offshore markets.

India has 10 unicorns such as retailers Flipkart and Snapdeal. Flipkart is the eleventh-ranked unicorn in the world at $11.6 billion. It is the biggest e-tailer in India, beating out U.S. giant Amazon.com. And then there is Alibaba and its rival Tencent.[1]

Secure Messaging services like Telegram, ProtonMail, Line and Signal are building virtual ecosystems with their tools

.

Opportunity

What does this mean for advertisers? For one, better targeting of their audience. More timely and relevant metrics. Personalized marketing, branding and a more granular understanding of behavior. But it comes at the expense of consumer privacy. So, the question is, "why do we need data aggregators that capture web behavior information without rewarding the consumer?" The goal would be to:

- **Reward a consumer sharing their information, preserving their privacy.**
- **Improve response rates for advertisers in a P2P Web.**
- **Provide accountability for use of data as a standard Web 3.0 practice.**

Privacy - GDPR

GDPR is the most important change in data privacy regulation in 20 years[2]. Its aim is to protect all EU citizens from privacy and data breaches. Its hidden aim is to move away from an opt-out to an opt-in model of data acquisition and sharing. This spells trouble for any application that uses services supported by the Internet. For data aggregators, it requires a more formal disclosure of how this information is reused by third parties, especially advertisers.

The current litigation by the EU against mega-firms like Google and Facebook could result in fines of several billion euros and these companies are prepared to pay, ostensibly to preserve their profitable targeting tools for search engines and social media. As direct data aggregators (capturing user metrics) their entire profit model is based on leveraging data gathered from site visits, social networks and user behavior across partner data systems. It's a giant industry.

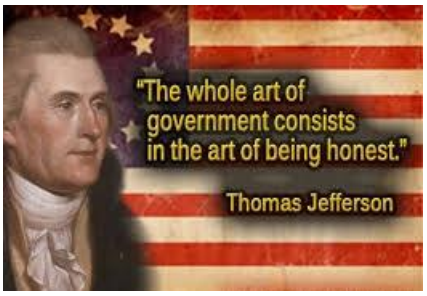
In the US and other countries, GDPR guidelines are beginning to appear but are stuck in regulatory mazes. In Japan, the US, India and South America, legal standards widely differ and political awareness of what is at stake and are lagging the revolution in Web technology, led by the global adoption of mobile smartphones as the preferred point of access.

South Korea seems to be more adept at embracing the momentum of privacy and e-commerce working together. Its Personal Information Protection (PIPA) was enacted in 2011 and is one of the world's strictest privacy laws. Like the GDPR, it protects privacy rights from the perspective of the data owner and applies to most organizations and government entities.[3]

Privacy – US

In 1999, Scott McNealy, CEO of Sun Microsystems opined that "You have zero privacy anyway, get over it." That was 20 years ago, and he was right. Sun was a member of the Online Privacy Alliance, an industry coalition to head off government regulation of online consumer privacy in favor of an industry self-regulation. His business cohort Eric Schmidt followed Scott's advice when he took over Google. OPA is now history.

In the US the Federal Communications Commission has been usurped by lobbyists of the vendors that move data through the Internet: network backbone carriers like AT&T and Verizon, data distributors like Akamai, cable vendors like Frontier and Spectrum, as well as data aggregators. Opt-in concept of GDPR is alien to the FCC idea of privacy and often referred to as not in the Constitution's Bill of Rights, enacted in the 18th century. Go figure.



Privacy – China

China is a mobile-centric country where most economic transactions occur online. Cash is a rapidly declining option for payments. The smartphone is the both the means to connect and transact. While it seems that this is a positive direction, China's ulterior motive is to monitor activities of citizens with technology. Facial recognition, GPS, monitoring user generated media, regulating social connections, incarceration of religious and ethnic groups, suppressing free speech; these are some of the things China is embedding into their Internet world.

Chinese users of the web are embracing the use of virtual private networks (VPNs) to get around this digital wall. Its impossible to regulate which apps users can download to their phone. The government can only control the Internet traffic to hosting sites for apps. But this won't work in a P2P network architecture. As app host services become decentralized, then blocking IP addresses (blacklisting) becomes futile. Still, monitoring data traffic and what it represents only works if it is not encrypted. Yet that is the direction of a secure Web 3.0.

Monetization in Advertising

From the early days of a commercially focused Internet, how to pay for the infrastructure, bandwidth and storage became an issue. What emerged was the idea that people would pay for access to the Web (telecom, cable, mobile contracts) but the ecosystem would fund its operations through advertising.

Web publishing sites would offer limited space for advertisers to deliver their message: through ad inventory delivered as content on a visited page, through popups powered by JavaScript and HTML5, with pre and post video ads for rich media, plus ads directed to mobile apps. In essence, the advertisers create revenue to web enabled services that offset their costs. Today this business model continues especially in gaming.

The emergence of social media changed the landscape. **Facebook** offers a rich place for users to share their profiles and activities along with a place to store generated content and share it using collaboration tools. Links, photos, articles, blogs. The key point is that their framework acts as a place for each subscriber to act as their own "website" albeit using Facebook tools. All for free. In return the site provides third parties the ability to target subscribers using more sophisticated methods for monetizing their profile data.

Is this wrong? Not from the aspect of providing a free service that allows a vendor to create revenue to support the service. It's a bit of a Faustian bargain. Here something for free but in return I can take your data and offer it as a way of providing prospects to advertisers.

But, the difference between Web publishing sites that may not know who you are or have limited data if you subscribe to their site, is radically different from a social media community. In the latter, what you post or view as your behavior lasts longer than an online session. It becomes fodder for updating your profile and creates a value measure that dictates how someone like Facebook charges third parties. Your data, their revenue, no privacy.

In a similar way, **Google** provides tools for web publishers to analyze site traffic in terms of a viewer's physical location, recency and frequency of visits and a whole host of metrics powered by cookies and site beacons. They become the traffic cop for a web publisher. In a more insidious way, internet service providers (ISP) can monitor traffic using IP addresses, device unique fingerprints and duration of sessions, like the old telephone networks did before the advent of the Internet. Along with the optical fiber network operators, they know most of what you do on the Web although they are not adept at monetizing this information or are prevented from doing so. Until now.

Privacy Economy

In previous articles on privacy we proposed something now considered as a positive step in monetization namely, the PRIVACY ECONOMY[4]. The premise is that new algorithms within AI science can replace the need for data aggregators who hold personal data in their analytic engines to help behavioral targeting for others such as online advertisers. We call this ***edge-AI*.**

> "Cloud computing agility is great – but it isn't enough. Massive centralization, economies of scale, self-service and full automation get us most of the way – but it doesn't overcome physics – the weight of data, the speed of light. As people interact with their digitally-assisted realities in real-time, waiting on a data center miles away isn't going to work. **Latency matters**." [5] – *Thomas Bittman Gartner*

Edge-AI

What is it? Breaking it down to components, **Edge** refers to an endpoint, your smartphone or computer as one example. An IoT device is another. It supports the ability to "compute at the edge. Instead of communicating to a central point to exhibit "smartness", algorithms can do the job at the endpoint. The **AI** part is a bit more difficult to define, due to how practitioners and media describe it. Simply put, it is a new science that focuses more on machine learning than human directed actions. Here are a few examples.

- Detecting a new object in an image without training – *cancer research*.
- Translation across languages – *Japanese – Chinese – German*.
- Finding exceptions to any process using pattern recognition – *defects & anomalies.*
- Discerning predictable order in a chaotic system - *weather*.
- Classification of physical phenomena – *facial & voice recognition*
- Extracting meaning from communications – *semantics & ontology*.

AI is a toolbox of scientific data methods that augment and sometimes replace procedures that require human intervention. Whether its is neural networks, classic statistical methods, or specific techniques, the goal of AI is to provide some form of intelligence with unsupervised learning. The ideal goal, as formulated by Alan Turing, is a system that *learns and adapts*.[6]

The two words combined imply that "intelligence" is no longer dependent on a centralized system of resources. But it's not easy coordinating edge enabled resources to create intelligence. The proliferation of computing power to endpoints is already there. 5 years ago, connecting your doorbell to a camera to your phone was just a concept. Not anymore. Connecting IoT sensors into ad-hoc networks where the latter can adapt to what data is flowing in real-time is a reality. Earthquakes, Tsunamis, Fires all are monitored in many areas.

Privacy & Edge AI

In the past, those who belonged to a village, knew what was going on and by whom. There were fewer people and word of mouth worked well. As villages became towns, then cities and states, it was cumbersome to communicate what was important. Local governments were used to establish a social contract with rules and permissions. The goal was to ensure stability and growth of its constituents. Each step however, placed a price on privacy – what a constituent wanted to make public. Often, political abuse made this price greater than what one was willing to pay. Enter regulation. Sometimes privacy was codified into law such as with religious orders. The important take-away is that humans value privacy but are willing to exchange some of it for incentives.

In a digital world the above is just as valid as it was 1000 years ago. We accept free services with the expectation that our data is used for the "good" of the commons. It's a nominal or base price for using services. The problem is that the concept of good has been hijacked by vendors through their EULA legalese, their customer service agreements, their public relations messages and now through their legal responses to groups that want to change the terms of what privacy means. It's adversarial and that is not a good path.

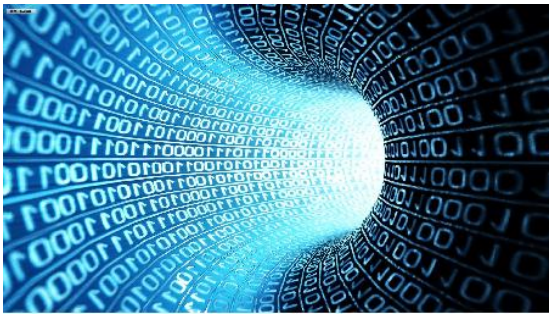
We are now at a point that we exist physically and digitally. The former is under our control. The other is not. Our digital personas are created by third parties and used to benefit or to economically punish us. Your FICO score tells people if you are worth something. Your health records impact your costs for survival. Your social comments act as a filter to include or exclude you from groups of similar minded people. Every action has an economic measure.

As humans, most want to belong. Living in isolation is an exception. To belong implies that we have something to offer to others. Artists choose other artists or people that like what they do. Same goes for authors or people engaged in social activities. Even criminals abhor a vacuum. In a digital world we face the prospect that technology can decide how we belong and what is our value to the group. This is the challenge. Advertising is just one trivial example. What's missing is a radical way for creating value as incentives.

AI Digital Profiles

Our digital profiles are created by third parties. They have many holes and tend to be outdated and incorrect. They are more static than dynamic and not easily exchanged. It's a world of data lakes and silos, each party aware that someone will hack this information and reduce its value. The main problem is data centralization. It affords a rich target to anyone that wants to steal the data.

Now imagine that you have multiple profiles or personalities. Each reflects a different category of interest. What if you could convert these profiles into a value-based data model. One for finance, another for healthcare, sub profiles for interests and product choices, what you consume or order, where you travel, your habits, your network of friends, opinions, information you pay for, groups you belong to, even political affiliations. This is a multi-dimensional view of you as a person. In the world of advertising this would be a gold mine to exploit and I emphasize the word "exploit". Its obvious that a perfect representation of your digital personality is worth something. The question is how to pass that worth back to the person - in a fungible way that can be used in future transactions. And this data is stored at the edge with you, not a third party.



So how does an external actor take advantage of your profile without invading your privacy as is want with data aggregators? We discussed this approach in our Privacy Economy papers (see footnotes.) While we have talked about securely keeping profile data at the edge, data science is based on creating intelligence from data by feeding a system from many data points.

AI is perfectly suited for collecting many data points to create a statistical predictive model. The data method does not care who you are (your profile). It wants to use your data to make the model reliable. To this end, every data point can be anonymous as it feeds the inference engine that can predict an outcome, not who is part of the outcome.

In this design, your profiles are private to you. These are calculated at the edge (example smartphone) to create multi-dimensional values (if you wish, scores.) Periodically, your MD-values are anonymously transmitted to a central AI engine that updates a global representation of values from all participants. In return, the global MD-values are transmitted back to you as a two-level filter that allows you to see where you exist inside the model. You can elect to be targeted on your unique profile or the group profile. (How we maintain anonymity between the central AI data and you, is described in the Privacy Economy article.)

AI Profile Monetization

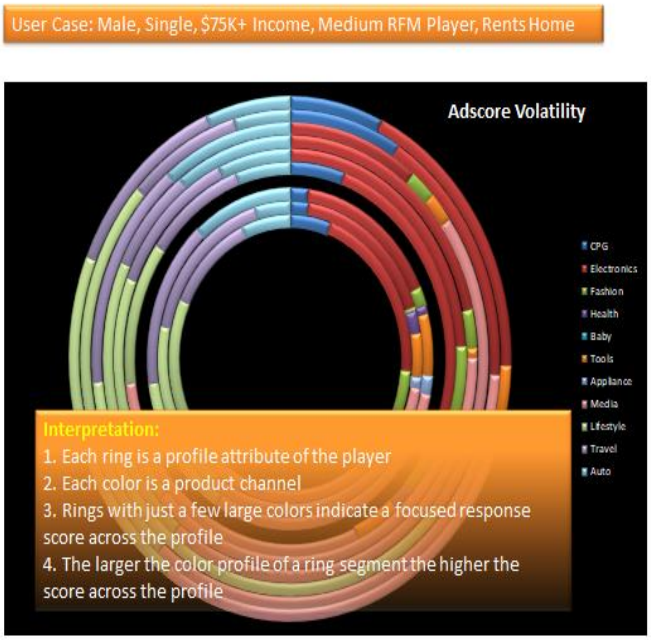
Earlier, we talked about a transformation that could occur for advertisers in a P2P decentralized Internet. Web 3.0 assumes that we have a digital identity that can be accessed using IP addressing. It requires migrating from IPV4 addresses to IPV6. Most people don't realize that this has already happened with smartphones, laptops, server resources and more recently, IoT devices.[7] The ability to use IPV6 as a means to preserve anonymity through anonymous agents is a perfect way to allow edge-AI profiles to benefit online commerce and advertisers.

Staying with the advertiser model, a firm wants to promote to a larger audience that share certain characteristics. The edge-AI model provides an anonymous unique signature to each participant. The centralized anonymous data store dynamically updates the aggregate MD-values of participants. These define the potential audience in any respective category. The advertiser sets a quota (impressions) that they are willing to deliver for a price to any participant with a score value that is within the range of their target group. Note, the price is to the participant, not to some data aggregator that performs the ad delivery.

Edge-AI Process

Here is where it gets interesting. The promotion can be delivered directly to the participant using different channels (text, email, video) or pushed through an ad network to a mobile app or a traditional web publishing site. (It could easily extend to other media forms like cable). In all cases, the advertiser allocates a fixed reward to the participant or viewer. Monetizing the value is another discussion but it is very flexible given SEO and SMO strategies.

The participant is willing to provide a measure of their online activity, in the form of MD-values that can be used to better target the message to the right audience. The edge-AI process preserves the granular detail of activity and only anonymously shares it with a central AI inference engine; that data sent is preprocessed for immediate update to the central data store.



Any advertiser can query the central data store as to audience potential and commit to paying for access using standard CPM/CPC/CPR pricing. All this conforms to the traditional web publishing site model except the message flows directly to a participant using multiple channels.

What's important is that the participant or viewer is rewarded directly by the advertiser from their marketing budget, that is in some level redirected from paying a data aggregator who is not interested in the privacy of the participant. Edge-AI meets the needs of participants who want privacy but seek an incentive to respond, and advertisers who want better targeting for the same marketing cost.

Summary

The environment for advertising support for Web services is changing. The old Web publishing model no longer works. The Internet is fast moving to a distributed and decentralized P2P network of services. Mobile computing and IoT devices are driving intelligence to the edge, away from centralized analysis. Humans are increasingly mobile-centered and new hardware supports more complex operations on their devices.

The world of data aggregators is rapidly shifting from existing targeting methods towards AI enabled systems. However, they are reluctant to have users in control of their data as it minimizes their value.

In a P2P Internet, users have the ability to create virtual associations that don't depend on the likes of Facebook of Google. As this movement continues, "smartness" moves to the edge user. Emerging tools allow performing AI operations at the edge with/without central AI systems. This a natural evolution of how the Internet was first described as - a P2P network. With IPV6, there are no limits to this growth.

Advertisers are spending their online budgets with third party data aggregators, some of which have not complied with self-governing privacy guidelines and are now in dispute with government regulators.

New technologies that protect privacy and improve targeting are replacing the brute force central storage methods employed by the likes of Google, LinkedIn and Facebook. As data moves to the edge, they cannot leverage their analytic value with existing centralized tools. They are committed to adopting AI to improve their offerings but without privacy controls, it will fail.

> "Advertising is a distraction, not part of my experience. That's why I ignore it" – GenX

Finally, users need to be monetized or rewarded. This means that a portion or all of the advertising revenue that companies spend to reach prospects can be redirected in novel ways to them. Data aggregators no longer need to be prime recipients of user information.

A P2P decentralized Web will make this happen, regardless of the future of technology defining Web 3.0. We are at the moment of disrupting the existing way that monetization of Internet services work. Users will now be part of the monetary equation and rewarded by the right message at the right time at the right place.

Author

Andre Szykier lives in Bermuda Dunes California, is a mathematician and founder of **UbiVault**, a leader in security solutions for a decentralized Web. He is also the CTO of a cryptocurrency ATM network – **BlockchainBTM**

Andre andre@ubivault.com

References

[1] Nikkei 2018
[2] EUGPDR 2019
[3] South Korea PIPA 2001
[4] Privacy Economy 1, 2.
[5] The Edge Will Eat the Cloud - Bittman
[6] Turing Test
[77] IPV6 Trends