"Digital Footprint": An Impression Left

Neha Kishore[1], Priya Raina[2]
Chitkara University, Himachal Pradesh
(E-mail: [1]neha.kishore@chitkarauniversity.edu.in
[2]priya.raina@chitkarauniversity.edu.in)

Abstract—The paper at hand deals with the impressions or traces that are left to a variety of customers when the user surfs online. The information left is used by advertising companies which target their customers with advertisements according to their web searches online. We have discussed the origin of these traces, the levels of concerned customers, key findings and surveys, advertising formats and an area which can be explored to a new level i.e. SCI social and community intelligence which deals with sensor based activity. The information is collected through sensors that are present in our mobile phones and other electronic devices used for searching on the internet. We have discussed various precautionary measures to be taken while posting anything, surfing or putting your personal information online as no one wants to lose his/her identity.

Keywords—Behavioral Targeting; Cyber Vetting; Digital Footprint; Sensor Based Activity; Virtual Shadow;

I. INTRODUCTION

A trace, trail or "footprint" that people leave behind while browsing on the internet is called a Digital Footprint or Digital Shadow. It is an online portfolio of who we are and what we do. The information transmitted online can be in the form of email attachments, recent web searches, forum registration, videos or digital images uploaded by an individual etc.[1]. All these traces are available to others online and are used to target marketing, to personalize, to promote socialization and to establish reputation. Wherever you travel throughout the social media, you're leaving digital breadcrumbs for others to follow.

This paper explains the significance of Digital Footprints - their origin, the threats they pose, and the opportunities they present. It also investigates preventive and curative measures for managing them. Section 2 discusses the preliminaries - how the footprints originate and their classification. Section 3 presents the key issues that arise as a result of digital footprints as well as looks at the developments due to them. It talks about how the advertising industries are minting money by just one click of a customer. The contribution of Facebook and Google has been portrayed in the advertising industry. We have discussed the concept of sensor based activity i.e. extracting social and community intelligence from sensors that are embedded in our devices that have access to the internet. We have taken into account their major characteristics and application area. The last section is about Management of Digital Footprints. It begins with an assessment of users, classifying them according to their levels of concern. It is followed by findings of a survey that gives us an insight into the average size of digital footprints of the users and levels of awareness about controlling the size of Digital Footprints. We have thrown some light on the steps that one should take while being online to "clean" his/her digital footprints. Finally, we study the merits of the idea of making the Data-Owner the Data Controller. Our main aim was to make each individual aware of the activities that are taking place online and we have discussed some measures that can help in providing least information online and be less visible to unknown people.

II. BACKGROUND

A. Origin of your Virtual Shadow

In the virtual world, as in the real world, we leave behind a trail of our activities - Digital footprints. The term is generally used in the context of an individual, but business, corporations, organizations too may have a virtual shadow [2]. They are visible in as trivial things as the frequency and amount of time spent by a user on a website. Digital footprint comprises of digital data and metadata (data about data), left behind due to mobile devices, Internet/Web Browsing and T.V. interactions. Thus, they can "brand" the user implicitly[3].

A unique dimension to the digital footprint is added through mobiles since they provide new content through web searches, location etc. They also set the social context for the digital footprint. Mobile devices have beaten other sources of digital footprints, both in terms of quantity and quality. Figure 1 [4] shows a visual, comparing the contribution of the three sources to the size of Digital Footprint. However, it must be noted that with the emergence of various "smart" devices and rise of concepts like Internet of Things, it is becoming difficult to distinguish between the data streams - they have become more integrated and to study their individual contribution is getting somewhat redundant.

B. Classification of Digital Footprint

Digital footprints can be classified as active or passive. Active footprints are left behind by user knowingly, e.g., personal information on social media platforms. Passive footprints are unwittingly left behind by the user e.g., while web surfing. Table 1 differentiates between the two.

III. USING DIGITAL FOOTPRINTS: KEY ISSUES AND DEVELOPMENTS

Digital footprints are left behind by users, whether inadvertently or deliberately:

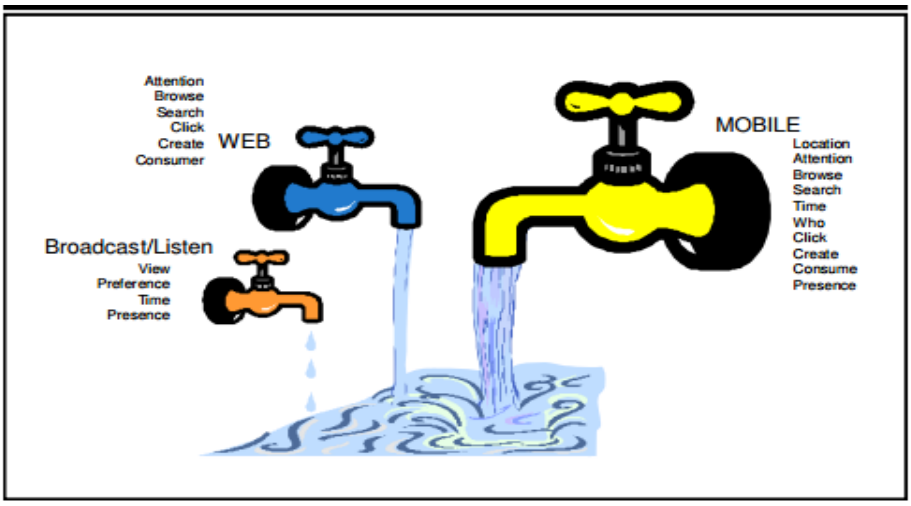

FIGURE 1. INFORMATION INTERPRETED FROM WEB DATA, MOBILE AND TV [4]

- As a result of a user's stored cookies and web-browsing information is left behind [2,5].
- Presence and activities on various Social Media Platforms may be visible not only to the target audience but also to unintended strangers.
- Various sensors within devices, particularly mobile devices, may lead to collection and analysis of data without the user being aware of it.

Once your information is online, it could be there forever and could be tracked by others, thus adding to the vulnerability of the user. Thus, users must be very conscious of their digital footprint [6]. However, data suggests that despite the precautionary measures by 86% of adult internet users, 11% had reports of stolen vital information like Social Security numbers, banking details and 21% had been victims of hacking attacks and identity theft [7]. Thus, the real issue is the amount of information available online.

Another issue that has emerged is the debate about privacy of individuals and groups on one hand and security on the other. This issue has gained significance since the revelation of Prism surveillance program, as also in the wake of effective use of social media by terror organizations. According to Esther Dyson, the only likely solution to protect the user from third party surveillance is to have more granular control over our data. She believes that the challenges of the current online environment are not fully captured by privacy[8] and that there is a need to control data rather than just talking about privacy. Similar thoughts are shared by Scott McNealy of Sun Microsystems. He terms consumer privacy as a red herring as we have zero privacy and we should all get over it". This view has gathered credence after 9/11[4].

Yet another issue pertains to the fact that the collected personal data is typically not available for the individuals themselves. The use of this data is restricted by the functions provided by the owners of these centralized points of collection, usually a corporate giant like Google, Apple, and Amazon etc. In fact, the data may not even be used in the individual's best interests, as it is controlled by another entity with sometimes conflicting interests. Furthermore, the user may be unaware of what data is being collected and stored. Finally, due to presence of several players, the footprint is often scattered, with data being stored in proprietary silos with little communication. In order to realize the benefits of data, users are forced/incentivized to use the services of a single vendor, something detrimental market competition. However, even in this single-vendor scenario, personal data is ultimately controlled by the vendor, not the user [9-11].

A. Third Party Surveillance

It is very easy for other parties to engage in stalking and third party surveillance by means of data gathering, active or passive, combined with the services of a simple search engine. Interested parties use digital footprints for a wide range of activities including Cyber-Vetting[12] (Background check for job applicants), surveillance by law enforcement agencies [13] and so on. Such usage allows digitally tracing data including social groups, behaviors, individual interests and location.

B. Behavioral Targeting

Behavioral targeting uses information and search history related data collected from an individual's web-browsing behavior to display advertisements. It refers to a range of methods and techniques adopted by online website publishers and advertisers so as to increase the effectiveness of advertising[15]. Recent developments in Big Data mining and tracking technologies have led to more sophisticated tools for Behavioral Targeting, which can capture and analyze detailed data about a user's action and online behavior. For example, a study by the Wall Street Journal [16] found that the nation's top 50 Web sites install, on an average, 64 pieces of tracking technology, usually without any notification to users[17].

TABLE 1. CLASSIFICATION OF DIGITAL FOOTPRINT

Passive Digital Footprint	**Active Digital Footprint**
When data is collected without client knowledge	Data released by user
Offline use is stored in files as they are created.	Created online when logging on, editing and posting. Created offline through a key logger.
Web browsing leaves a digital shadow gathered by website hits and cookies by recording IP address, time of access etc. This information is available to and used by marketers, researchers, law enforcement agencies (no probable cause required)	Social media networks leaves behind user's personal data that includes personal interest trends, social affiliations, behaviors and location.
Some proxy servers can even collect every keystroke made by the user.	Personal privacy settings can be compromised by applications within social media sites[14]

Advertising Formats

Different advertisement formats found on the internet have been improvised by using Behavioral Targeting -

- A text Webpage: Banner advertisements associated with text WebPages (e.g. from Dictionary.com or MSN) may be selected to reflect a user's interest.
- A video Advertisement: A "Pre-roll" i.e. a video advertisement that appears before a requested video starts.
- Overlay Ad: Advertisements appearing near the bottom of the window showing requested videos to the user (e.g. on YouTube or ESPN3) may be tailored based on user's interest.
- In-App Ads: Almost all applications on Smartphones flash advertisements to the user.
- Internet based advertisements are auctioned off on the click-through rates basis and user's interests are derived from their on-line browsing behavior.

Facebook Exchange (FBX)

FBX was launched in September 2012. FBX, put in the simplest terms, is an inventory of ad space that outlets can bid on in real-time, and retarget to their respective audiences[18]. Tracking cookies is installed on your browser so that the advertisers can look at what you're surfing on the internet. These advertisers who buy ads on Facebook, target you with ads they just bought and are customized according to your search history. Google has been doing this since years but FBX launch is still noteworthy since it has the potential to change advertising. Facebook was home to 25% of the Internet's inventory of display ads. Facebook has teamed up with Digital Advertising Alliance and has come up with an Ad choice logo for users to opt out. Now when you hover over a targeted ad in your sidebar, you can hide that particular ad by clicking on 'X' pop up in top right corner[18].

There are many ways by which websites are collecting one's personal information. The main aim is that everyone should be aware of the amount of data available about them online. Some sites collect data which is used in a productive manner but for some people even that information can be objectionable. This is discussed in our next topic that how sensors in our devices collect information and what should be done to reduce your digital footprints.

C. Sensor Based Activity: Extracting SCI from Digital Footprints

SCI (Social and community intelligence) is an emerging area in the wake of ubiquitous and all-pervasive, sensor equipped mobile devices powered by ICT[19]. According to the Pew Internet Project's research related to mobile technology (2014) 90% of adults have a cell phone[20]. Millions of vehicles are equipped with a GPS system nowadays and the mobile location technology is expected to grow at a higher rate. Twitter, Facebook, MySpace and LinkedIn record information related to one's relationship and preferences.

Sensors common to most mobile phones are:

- RFID (Radio Frequency Identification)[21]: For example if each household item is tagged with an RFID and a person wears a watch like RFID reader, all his actions can be tracked [22].
- GPS (Global Position System): It is used for location based activity recognition [23].

Characteristics of SCI Research

[19] identifies the following characteristics-

- Infrastructure: Infrastructure for integration of data from variety of devices with different (possibly incompatible) configurations. Real time data sensing and inference is a key feature.
- Data Sources: They are heterogeneous and multimodal. Three main data sources are about
  a) an individual through mobile/wearable sensor data
  b) The environment through infrastructure-bound sensor data.
  c) Social data about individual.

- Technology: Data mining, artificial intelligence and machine learning are the core technologies for SCI.
- Application: Innovative services like public safety, healthcare, are enabled.

Major Application Area

It is an upcoming field. So far following applications have been identified-

- Friend Recommendation: Two person's interests, behavioral data and profiles are matched by a serendipity system and then it alerts that someone might be interested in him/her.
- City Resource Management: Bike community in Los Angeles is improved by a Biketastic project which enables the bikers to plan a best possible route.

This is an area which will grow with time since everyone has a mobile phone or any electronic device to surf the internet but no one is aware that how their activity is recorded. Everyone has used the embedded sensors in their phones but due to lack of proper information one tends to ignore the fact that each and every step online is recorded if proper measures are not taken[24].

IV. MANAGING DIGITAL FOOTPRINTS

A. *Levels of Concern*

Being aware of your digital footprints is a necessity because it may carry the record of the user's entire life, right from their first baby photographs to the activities they currently participate in. Users can be divided into four groups depending on their level of concern about their online information, represented in Figure 2 [7]:

- Unfazed and Inactive group (43%) is the largest of the four groups. They do not worry about their personal information and do not take any steps to limit the amount of information.
- Concerned and careful group (21%) takes care about the information online and take precautionary measures. One out of five online adults fall into this category.
- Worried by the Wayside group (18%) does not take any step to limit their information even when they are anxious about what is available online about them.
- Confined Creative group (17%) does not worry about their information online and upload content actively, but still take steps to limit their personal information.

Thus, level of concern for a person is largely determined by how important the information is for them and how highly they value privacy and anonymity.

B. *General Trends*

Findings of a survey conducted by Pew Research Centre [7] in 2013 can give us an insight into the size of digital footprint of users and the measures used by them to manage it. This section summarizes those findings. However, the findings must not be generalized since the survey had been conducted in USA, a country where digital literacy is quite high.

Users report that a wide range of their personal information is available online, but feel strongly about controlling who has access to certain kinds of behavioral data and communication content. But still people are becoming aware and taking measures to safeguard their digital footprint.

Figure 3 represents the information revealed by most people on internet platforms in descending order. Figure 4 shows the variety of ways that internet users have tried to avoid being observed online [7].

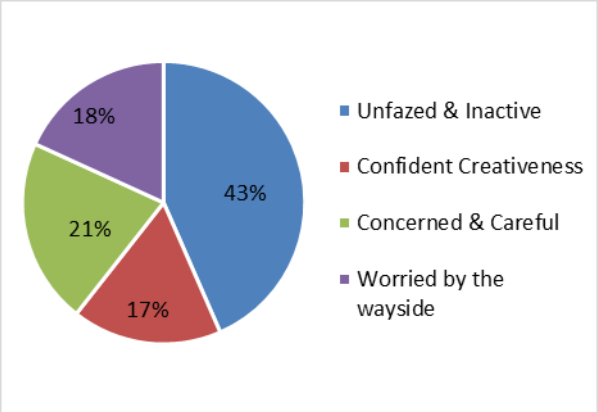

FIGURE 2. LEVELS OF CONCERN FOR DIFFERENT GROUPS OF PEOPLE

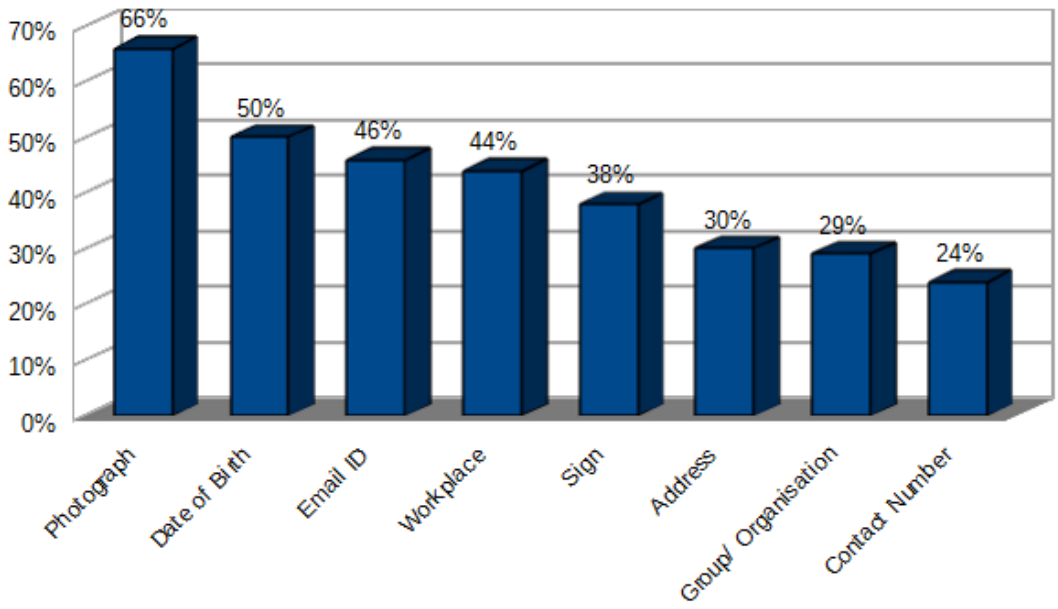

FIGURE 3. PERSONAL INFORMATION ONLINE

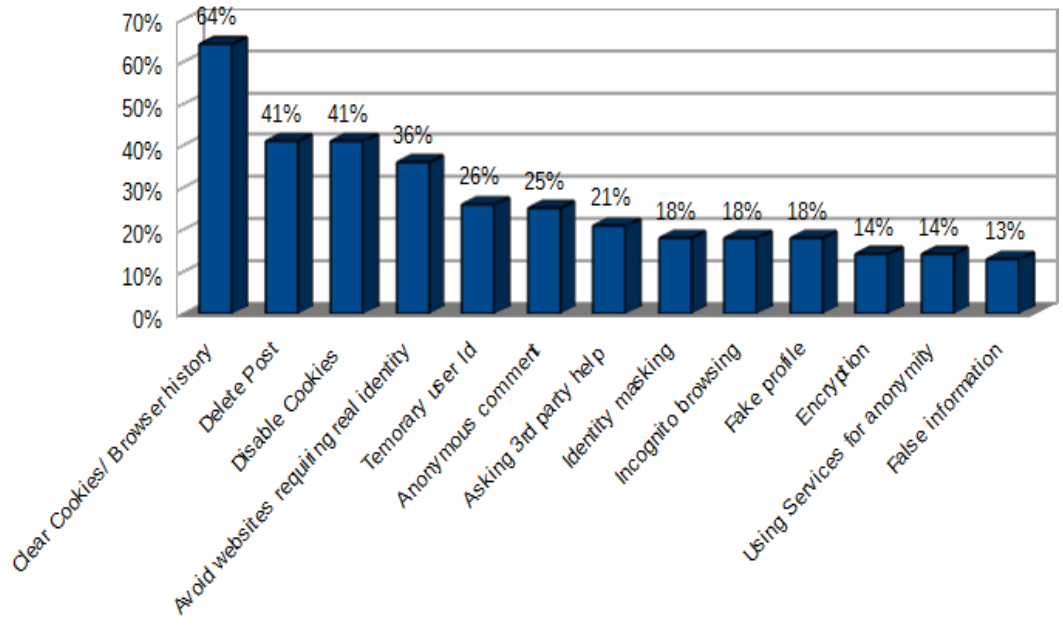

FIGURE 4. STRATEGIES FOR FOOTPRINT MANAGEMENT

Source: Pew Research Center's Internet & American Life Project Omnibus Survey, conducted July 11-14, 2013, on landline and cell phones. N=792 for internet users and Smartphone owners. Interviews were conducted in English on landline and cell phones. The margin of error on the sample is +/- 3.8 percentage points [7]

C. Cleaning Digital Shadow

With proliferation of PCs, Smartphones and ease of Internet access, people's digital footprints are floating around in the cyberspace, growing in size- thanks to multiplicity of accounts, for multiple ends, like social media, e-commerce etc. They can be accessible to anyone willing to look for your information and this can be a disturbing factor. Campbell-Young says it is hard to erase the trails that we accidently leave behind. Two major factors i.e. growth in identity theft and rise in corporate tracking are the main reason why deleting or controlling digital footprints is a critical task. He says Google has been correctly held answerable for collecting our data and is one of the biggest culprits[25]. The ads that pop up on your mail are customized for you, considering your past searches on the Google search engine.

Erasing your digital shadow completely is not always possible, but you can certainly come close if you follow these steps:

- Do a Comprehensive Self Search – See what information is available about you online by looking yourself up using a search engine. If it is potentially damaging to your reputation, get it removed. Remove tags from pictures, as well as flagging information for removal[26]. This activity must be done regularly. Setting up search alerts about your name may help, especially for uncommon names.
- Use Maximum Privacy: Utilize the Privacy Settings to the fullest. However, before agreeing to any privacy policy, thoroughly go through it and only then agree to the content written. Scrutinize each bit of information you have posted about yourself to make sure your digital footprint is as clean as possible[26].
- Secure your Name: Create a new email account with a more unlikely username with a free email provider. (Ex.: m24v_!96$7lkp@whatever.com)[26].
- Tackling undeletable accounts: Some sites don't allow you to delete your account completely. They ask you to "deactivate" your account with your information retained in their system. If there is a real reason for removal (such as witness protection), contact the site's owner or engineers and seek help. But if no help is available, delete as much information as possible and don't hesitate to provide false information in compulsory columns[26].
- Clean up your computer: Remove internet history, cookies, etc. from your mobile phones or PC's. Blocking cookies isn't viable for most of us as cookies are essential for using many sites, including Amazon and Facebook. However, "Do Not Track" feature on most modern browsers can come to our rescue.
- Think before you post anything online[27] : Always give a second thought before posting anything on social networking sites or passing a comment to someone. Ask yourself the following:
  a) Who might be able to read this?
  b) Could someone misinterpret what I'm saying?
  c) Am I posting in anger?
  d) Am I showing a bad side of myself?
  e) Could someone feel disrespected?
  f) Am I revealing too much about myself?

To be safe online you need to take these measures and try to put least personal information online. Since these traces are very hard to erase one should be aware of his/her activity online and should not disrespect anyone as one ill lawful activity can lead to strict actions.

D. Human Centric Control of Data

It is the idea that Personal Data should be controlled by users themselves, rather than by any third party. Two main approaches have been proposed to enable human-centric control of personal data:

- Centralizing the storage of the data: With this approach, the scattering of data is solved by providing individuals with a personal data storage service within which they accumulate data from various sources[28].
- Focus on managing the flows of data between data sources and data-users, rather than centralizing the data storage: In this case, the scattering of data with disparate third parties is solved by federation of data sources. The individual controls the uses of personal data by employing tools and infrastructure intended for managing permissions to access data[29].

V. CONCLUSION AND FUTURE WORK

This paper discussed all the major aspects of online activities and digital footprints. We discussed the main root cause of the traces left online i.e. online web searches, filling a form with personal information, web browsers and social networking sites. The main motive was to make the online users aware that their activities can haunt them one day so they should take precautionary measures like giving minimum personal information, thinking before making a remark, prefer do not track option and do not disrespect any one. Employees need clear guidelines for how to protect the reputation of the company as well as other colleagues in the social media world.

We discussed a new research area i.e. SCI and we'll be working on it about how one can be safe or send data without interference of a third party. We'll find new applications that can help identifying criminals from their online activities and find more measures to keep each individual's data more secure.

ACKNOWLEDGEMENT

The project is sponsored by SERB, Department of Science and Technology vide grant no. YSS/2015/001573.

REFERENCES

[1] "What is Digital Footprint? Webopedia Definition." [Online]. Available: https://www.webopedia.com/TERM/D/digital_footprint.htm2.

[2] K. Collins, "Monitoring digital footprints to prevent reputation damage and cyber attacks" [Online].

[3] M. Madden, S. Fox, A. Smith, and J. Vitak, "Digital Footprints:Online identity management and search in the age of transparency," Pew Internet Am. Life Proj., no. December, p. 50, 2007.

[4] T. Fish, "My Digital Footprint A two-sided digital business model where your privacy will be someone else's business!" [Online].

[5] S. Garfinkel and D. Cox, "Finding and Archiving the Internet Footprint," DTIC Document, 2009.

[6] "Get the facts- Digital Footprint." [Online]. Available: http://www.cybersmart.gov.au/Kids/Get%20the%20facts/Digital%20footprint.aspx. 2014

[7] L. Rainie, S. Kiesler, R. Kang, and M. Madden, "Anonymity, Privacy, and Security Online," Pew Res. Cent., vol. 5, p. 35, 2013.

[8] Ajit, "Social media marketing, Behavioural analysis and the transition from passive digital footprints to active digital footprints".[Online]. Available: http://www.opengardensblog.futuretext.com/archives/2008/11/social_media_ma.html.

[9] S. Abiteboul, B. André, and D. Kaplan, "Managing your digital life," Commun. ACM, vol. 58, no. 5, pp. 32–35, 2015.

[10] H. Haddadi, H. Howard, A. Chaudhry, J. Crowcroft, A. Madhavapeddy, and R. Mortier, "Personal Data: Thinking Inside the Box," arXiv Prepr. arXiv1501.04737, 2015.

[11] M. Sjöberg et al., "Digital me: Controlling and making sense of my digital footprint," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9961 LNCS, pp. 155–167, 2017.

[12] C. Dalgord, "Cybervetting: The Hiring Process in the Digital Age | Information Policy for Everyday Decisions - Wayne State University Blogs." [Online]. Available: https://blogs.wayne.edu/informationpolicy/2012/12/07/cybervetting-the-hiring-process-in-the-digital-age-2/.

[13] "Criminal Justice Information Services (CJIS) — FBI." [Online]. Available: https://www.fbi.gov/services/cjis.

[14] Follow your Digital Footprint. http://www2.huhs.org/library/pathfinders/footprint/footprints.html. 2014

[15] "Behavioral targeting - Wikipedia." s, 2017. https://en.wikipedia.org/wiki/Behavioral_targeting#cite_ref-Chen_2014_429.E2.80.93449_1-0.

[16] J. Angvin, J. Valentino-DeVries (2010) "Race is on to 'fingerprint'phones, PCs | Wall Street Journal " [Online]

[17] J. Chen and J. Stallaert, "An economic analysis of online advertising using behavioral targeting," Mis Q., vol. 38, no. 2, pp. 561–588, 2014.

[18] "If you don't want ads targeting you, read this story about Facebook FBX | Digital Trends." [Online]. Available: https://www.digitaltrends.com/social-media/what-is-this-facebook-fbx-nonsense-and-why-should-i-care/.

[19] D. Zhang, B. Guo, B. Li, and Z. Yu, "Extracting social and community intelligence from digital footprints: an emerging research area," in International Conference on Ubiquitous Intelligence and Computing, 2010, pp. 4–18.

[20] "Demographics of Mobile Device Ownership and Adoption in the United States | Pew Research Center." [Online]. Available: http://www.pewinternet.org/fact-sheet/mobile/.

[21] "How RFID Works|HowStuffWorks." [Online]. Available: https://electronics.howstuffworks.com/gadgets/high-tech-gadgets/rfid.htm.

[22] M. Philipose et al., "Inferring activites from interactions with objects," IEEE Pervasive Comput., vol. 3, no. 4, pp. 50–57, 2004.

[23] Ahmed El-Rabbany, Introduction to The Global Positioning System. Artech house, 2002.

[24] L. Wang, T. Gu, X. Tao, and J. Lu, "Sensor-based human activity recognition in a multi-user scenario," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2009, vol. 5859 LNCS, pp. 78–87.

[25] "How to remove your digital footprint from the Internet." [Online]. Available: http://www.allvoices.com/contributed-news/16082803-how-to-remove-your-digital-footprint-from-the-internet.

[26] "How to Delete Yourself from the Internet: 12 Steps (with Pictures)." [Online]. Available: https://www.wikihow.com/Delete-Yourself-from-the-Internet.

[27] "Think before you post online – Barking and Talking." [Online]. Available: https://spicylearning.wordpress.com/2010/11/07/think-before-you-post-online/.

[28] Y. A. De Montjoye, E. Shmueli, S. S. Wang, and A. S. Pentland, "OpenPDS: Protecting the privacy of metadata through SafeAnswers," PLoS One, vol. 9, no. 7, p. e98790, 2014.

[29] D. McAuley, R. Mortier, and J. Goulding, "The Dataware manifesto," in 2011 3rd International Conference on Communication Systems and Networks, COMSNETS 2011, 2011, pp. 1–6.

Dr. Neha Kishore, She has done her PhD in Computer Science and Engineering from Chitkara University, Himachal Pradesh, India in year 2015. Her area of research includes Parallel Computing and Information Security.

She is working as an Associate Professor in Chitkara University, H.P., India for last eight years. She has many research papers and poster presentations at International Journals/Conferences in her credits.

Dr. Kishore is a member of ACM, IAENG, Internet Society, UACEE, CSI. She has been certified as an ACM Ambassador.

Priya Raina, Junior Research Fellow, Chitkara University.

She did M.E (CSE) from Panjab University in 2016. Her research interests include Security and Cryptography.