I.J. Modern Education and Computer Science, **2015, 7, 36-47**
Published Online July 2015 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2015.07.05

Privacy in the age of Pervasive Internet and Big Data Analytics – Challenges and Opportunities

Saraswathi Punagin
Graduate Student, Dept. of CSE, PESIT BSC, Bengaluru, 560100, India
Email: sara.punagin@gmail.com

Arti Arya
Head, Dept. of MCA, PESIT BSC, Bengaluru, 560100, India
Email: artarya@pes.edu

***Abstract*—**In the age of pervasive internet where people are communicating, networking, buying, paying bills, managing their health and finances over the internet, where sensors and machines are tracking real-time information and communicating with each other, it is but natural that big data will be generated and analyzed for the purpose of "smart business" and "personalization". Today storage is no longer a bottleneck and the benefit of analysis outweighs the cost of making user profiling omnipresent. However, this brings with it several privacy challenges – risk of privacy disclosure without consent, unsolicited advertising, unwanted exposure of sensitive information and unwarranted attention by malicious interests. We survey privacy risks associated with personalization in Web Search, Social Networking, Healthcare, Mobility, Wearable Technology and Internet of Things. The article reviews current privacy challenges, existing privacy preserving solutions and their limitations. We conclude with a discussion on future work in user controlled privacy preservation and selective personalization, particularly in the domain of search engines.

***Index Terms*—**Privacy, Personalization, User Profiling, Pervasive Internet, Big Data Analytics, User Control.

I. Introduction

The internet is a fast growing phenomenon and the statistics released recently are a fitting testimony to this enormous growth. As per the 2014 Global Internet Report released by the Internet Society [34], the internet will have 3 billion users by 2015, the mobile broadband exceeded fixed users as of 2010, there were 1 billion internet hosts as of 2013, and developing countries have more than 50% share in the world's mobile broadband subscribers. In short, the internet is pervasive.

Use of internet has become inevitable and the convenience it brings gets attractive by the day. Whether it is buying things online, paying bills, connecting and communicating with friends and family or a simple web search, the internet has seeped deep into our everyday life.

While this kind of pervasive connectivity is a thing to applaud, it does bring with it several challenges; especially in the age of big data and extreme analytics. The digital footprint left by users is becoming worth its weight in gold because of the opportunities to store and mine this data for the benefit of smart business solutions. It is not only business that benefits from big data, but several domains including law and order, health care and research stand to gain from its virtues. No wonder it is now said that "Data is the new oil".

However it brings to light several privacy challenges. Data Privacy is defined as "the freedom from unauthorized intrusion" [4]. Internet users are usually unaware that they are leaving a digital footprint. They are unsure if and what part of their personal data is visible or who has access to it. They are oblivious to the fact that their data is being used by third parties for personal gain. However more and more users are gradually becoming privacy-aware. In a research study conducted by the Pew Research Internet Project[1], 91% of Americans felt that consumers have lost control over how personal information is collected and used by companies. However 55% of them were willing to share their personal information in certain circumstances when doing so gave them access to free services [22]. Therefore it is becoming increasingly important to find a balance between privacy and utility of personal information. It is time for service providers to adopt privacy preserving solutions and embrace the "privacy-by-design" paradigm instead of making privacy an afterthought.

It is a well known fact that privacy preservation comes at the cost of utility. The more privacy preserved a data set is, lesser the information it provides. A novel approach would be to put the user in control of how much privacy he is ready to sacrifice in order to get better utility. Also, as data collection and usage mining becomes transparent and users understand what data is being mined and how it is used, they may be willing to share their personal information with increased confidence.

This paper aims to review domains like Web Search, Social Networking, Health care, Mobility, Wearable Technology and Internet of Things and present privacy risks associated with them. We also present limitations in

[1] www.pewinternet.org

the current literature thereby listing open problems that need attention.

II. Related Work

Weber [48] surveyed the security and privacy challenges related to Internet of Things where solutions that address the privacy and security issues of IoT architecture and IoT data were studied. The study also suggested establishment of a legal framework that is flexible and takes into account the underlying technology and principles. Charu et al. [1] discussed IoT from a data analytics perspective and suggested various privacy preserving solutions for mining and managing IoT data. Zhang et al. [49] surveyed privacy and security challenges with respect to online social networks and addressed the utility vs. privacy preservation conflict and suggested opportunities to resolve these conflicts. Fung et al. in [50] discuss how Privacy Preserving Data Publishing (PPDP) is a unique problem and address privacy challenges in data publishing while summarizing different approaches to PPDP. Eran, Wang and Cranor [51] examine the privacy vs. personalization issues in various personalization based systems like social, behaviour and location based profiling.

All of the above work focused on privacy risks associated with specific domains where they weighed risks and discussed existing solutions to specific areas. There is no work which provides a holistic view of privacy risks and remedies in the age of extensive use of internet, collection of big data and state of the art analytics spanning all domains. Our survey attempts to fill this gap by surveying privacy/personalization challenges and opportunities across various domains like Web Search, Social Networking, Healthcare, Mobility, Wearable Technology and Internet of Things.

III. Big Data – Challenges And Opportunities

Data is generated from online transactions, social network interactions, machine to machine communication, web searches, health management, financial management and even entertainment [39]. Data is also no longer in a structured form but comes in all flavors including unstructured click streams, videos, pictures etc. Data is generated at great speed and in high volumes. Storing or transmitting such huge volumes of data is no longer a challenge with the advent of technological advancements in storage and networking. Analyzing such heaps of data in record times is also not a challenge any more because of the availability of newer algorithms and big data analytical framework. This is the world of big data analytics which brings with it several benefits as well as concerns.

A. Big data – Opportunities

1. Healthcare: Big data presents a lot of analytical and data mining opportunities in the healthcare domain. The discovery of inconceivable side effects by drug interactions is one such example. It was discovered through big data analytics that when Paxil and Pravachol are taken together they increase a patient's blood glucose to diabetic levels. For the one million users who were taking both of these drugs together, this discovery was a boon.[39] Researchers in South Africa discovered a correlation between therapeutic use of Vitamin B and the delay of progression to AIDS and death of HIV positive patients. In a region where therapies for people living with AIDS are expensive this was a critical discovery. [39] A more common example is of Google Flu Trends which makes predictions and locates flu outbreaks based on aggregated search queries. [39]

2. Mobility: Mobile devices are always on and are location aware with multiple sensors (cameras, GPS, microphones and Wi-Fi) [39]. This enables collecting of location based data and harvesting of the same for benefits like real-time traffic analysis, crowd handling, law and order and targeted location based marketing.

3. Smart grid: This includes sophisticated usage of smart grid data from utility companies to monitor and control electricity use, to predict energy demands, locate outages, and speed up repairs. Consumers also stand to benefit from this kind of analytics because they can manage their usage patterns to lower their consumption and go green.

4. Retail: Usage of big data analytics in retail includes and is not limited to real-time stock analysis (every store, every shelf), up selling and cross selling analytics (frequent item sets), targeted marketing, studying customer's in-store behavior to improve the store lay out and personalized recommendations.

5. Law Enforcement: Big data analytics has played an important role in law enforcement. Contributions include fraud detection and prevention, criminal behavior and suspect identification and forensic data analysis to name a few.

B. Big data - Challenges

1. Identity and Sensitive attribute disclosure: Even though data providers anonymize the data before publishing, such data is susceptible to external linkage attacks where an adversary is able to link publicly available data to the anonymized data and determine the identity of records. Narayan and Shmatikov demonstrated this in [27], when they de-anonymized Netflix's anonymized data set by linking it with user profile data from IMDB. Similarly Latanya Sweeny[36] was able to deanonymize a Massachusetts hospital's anonymized health records by linking the data set with publicly available voting list and determine the governor of Massachusetts' health problem (sensitive attribute disclosure).

2. Automated Recommendations: Big data analytics attempts to profile a user's worth by their click history, web searches, shopping habits and social interactions. A user is provided with recommendations that he may not agree with or necessarily be in need of. Recommendations at the least can be irritating. However they can result in a privacy breach if such recommendations were revealed on a user's social site without the user’s consent. Profiling also facilitates differential pricing and search discrimination on ecommerce websites which is controversial. Differential pricing or price discrimination, as defined by Mikians et al. in [26], is the ability of an ecommerce website to price a product on a per customer basis - using attributes of a user drawn from his user profile. The authors in [26] also suggest the existence of search discrimination, which is to steer a user to an appropriately priced set of products based on his profile attributes.

3. Predictive Analysis: While predictive analysis is invaluable when it comes to detecting and preventing crime or an epidemic, it can get out of hand if the same is applied to individual purchase habits. An infamous example is the New York Times story about retail giant Target’s “Pregnancy Prediction Score” based on a woman’s purchase habits. Duhigg in [8] sheds light on how Target identified about 25 products which when bought by their female customers according to a particular pattern, enabled Target to assign such shoppers a pregnancy prediction score. Most importantly it enabled them to estimate a due date, so Target could send them coupons timed to specific stages of pregnancy.

IV. General Privacy Measures

Before we delve into privacy it is important to understand the definitions of some of the privacy measures and the context in which they are used. Kifer et al. in [21] indicate that in privacy domain, a user is said to have the following three attributes:

- Identifiers: Attributes that uniquely identify a person. Example: social security number, user ID etc.
- Quasi-identifiers: A collection of attributes that is capable of identifying a person uniquely. Example: age, date of birth and zip code.
- Sensitive attributes: Attributes which the user intends to keep as private. Example: Sexual orientation, Disease information etc.

Table 1. Data before Generalization and Compression

User ID	Name	Ethnicity	Date of Birth	Gender	Zip Code	Health Condition
111	Alicia	Caucasian	09/27/1964	Female	19020	Incontinence
222	Bob	Caucasian	03/18/1963	Male	19022	Incontinence
333	Charlie	African American	04/18/1964	Male	19022	Incontinence
444	Emma	African American	09/30/1964	Female	19022	STD
555	Kevin	Asian	05/14/1961	Male	19022	Incontinence
666	Fang	Asian	09/15/1964	Female	19025	STD
Identifiers (Suppress them)			Quasi-identifiers (Generalize them)			Sensitive Attribute (Private)

One easy way of preserving privacy is to blend with the crowd. In the data or information context this means generalizing or anonymizing the data so as to not let one user’s record stand out. Generalization involves recoding an attribute value so it is no longer specific while Suppression refers to hiding an attribute value or not releasing it at all. Following are four approaches that have often been used to achieve generalization and suppression.

A. *k-anonymity*

As indicated by Sweeny in [36] a released data set is *k* anonymous if, within the data set, a record is indistinguishable with at least *k*-1 other records. *k*-anonymity ensures that even with external linkage attacks, each released record will relate to at least *k*-1 records and therefore provide no useful information to the attacker. Table 1 is the original data set while Table 2 represents the anonymized data set satisfying 3-anonymity.

Table 2. K-anonymized Data Set (k=3)

User ID Name Ethnicity	Date of Birth	Gender	Zip Code	Health Condition	Class
Suppressed	09/**/1964	Female	1902*	Incontinence	C2
	//196*	Male	19022	Incontinence	C1
	//196*	Male	19022	Incontinence	C1
	09/**/1964	Female	1902*	STD	C2
	//196*	Male	19022	Incontinence	C1
	09/**/1964	Female	1902*	STD	C2
Identifiers (Suppressed)	Quasi-identifiers (Generalized)			Sensitive Attribute (Private)	

Limitations of k-anonymity: There is a trade-off between data utility and privacy. A high value of *k* ensures better privacy preservation but reduces data utility. As indicated by Kifer et al. in [21], solutions with only *k*-anonymity are susceptible to sensitive attribute disclosure. For example, for the data set in Table 2, for an adversary knowing that Kevin is a male born in the 1960s is enough to determine that Kevin suffers from Incontinence. This is because there is no diversity in the sensitive attributes of the *k*-anonymized equivalence class C1. This brings us to the next measure in privacy preservation called *l*-diversity.

B. L -diversity

L-diversity as proposed by Kifer et al. in [21] ensures that the sensitive attributes in each equivalence class are diverse. A set of records in an equivalence class C is *l*-diverse if it contains at least *l* “well represented” values for each sensitive attribute. For example: For equivalence class C2, in Table 2, sensitive attribute of health condition is represented by (*l*=2) diverse values, namely, Incontinence and STD. So this class is 2-diverse.

Limitations of l-diversity: Li et al. discuss the limitations of *l*-diversity in [20]. For example, considering all six records in Table 2, we see that 2 out of 6 records i.e. 33.33% of them have the sensitive attribute value of STD. Now if we look at equivalence class C2 alone, 2 out of 3 records have the value of STD making it 66.67%. So the belief that an individual has STD increases in class C2. Therefore data sets where the distribution of a sensitive attribute in a particular class is very different from its distribution in the overall data set are susceptible to skewness attacks. This brings us to another privacy measure called *t*-closeness.

C. T-closeness

T-closeness as proposed by Li et al. in [20] considers the distance between the sensitive attribute distribution in each class, to its overall distribution. By definition, a set of records in an equivalence class C is *t*-close, if the distance between the distribution of a sensitive attribute A, in C, and its overall distribution in the record set is not greater than a threshold value of *t*.

When a record set is *t*-close, an adversary can learn very little individual-specific information irrespective of the amount of background knowledge he possesses about the record set. *t*-closeness resists the following attacks:

- *Skewness attack:* Since the within-group distribution of confidential attributes is the same as the distribution of those attributes for the entire data set, no skewness attack can occur.
- *Similarity attack:* Again, since the within-group distribution of confidential attributes mimics the distribution of those attributes over the entire data set, no semantic similarity can occur within a group that does not occur in the entire data set.

Limitations of t-closeness: Domingo-Ferrer et al. in [7] argue that while [20] discusses several ways to check *t*-closeness, it does not provide any computational procedure to ensure *t*-closeness is enforced. Domingo-Ferrer et al. in [7] also argue that enforcing *t*-closeness severely limits the amount of useful information that is released.

D. Differential privacy

Differential privacy is popular in mining of tabular data. As illustrated in Figure 1, it is an approach that adds randomness to statistical queries. Dwork in [9] suggests that the risk to someone's privacy should not increase as a result of participating in a statistical database. Task and Clifton in [37] suggest that "Differential privacy uses noise to obscure an individual's contribution to aggregate results and offers a strong mathematical guarantee that the individual's presence in the data set is hidden." By definition in [9], a randomized function K satisfies ε -differential privacy, if for all data sets D1 and D2, differing in at most one element, and any subset S of possible outcomes in Range (K),

$$P(K(D1) \in S) \leqslant \exp(\epsilon) \times P(K(D2) \in S) \qquad (1)$$

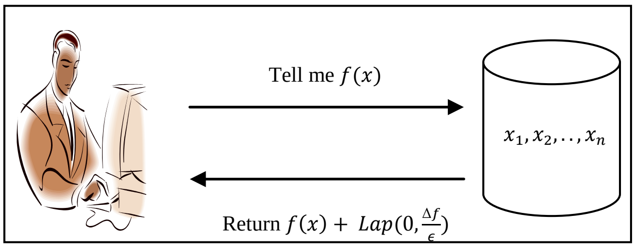

Fig.1. Differential Privacy

1. Achieving ϵ -differential privacy: ϵ -differential privacy is achieved by adding a random noise to the query results. Example: Let K be a count function = "Count the number of people older than 60." For data set of size = n people, Range (K) = {1, 2, 3 . . . n}. Sensitivity of K, i.e. K = 1 for any D1 and D2 differing in one element. The noise added to the query is related to the sensitivity of the query. Sensitivity is defined as:

$$\Delta\text{K}=\|K(D1) - K(D2)\| \qquad (2)$$

2. Limitations of differential privacy: Yang et al. in [45], list limitations of Differential Privacy as:

- The physical meaning of ϵ is not clear to real users making it hard for users to select the appropriate ε value to maximize the utility of the results provided by differential privacy mechanism.
- Most existing studies focus on querying static databases. Handling differential privacy protocols for arbitrarily updated databases is difficult and needs to be explored further.
- Current differential privacy techniques assume a central database with a single owner. It will be interesting to see how differential privacy fairs in distributed and multi-owner environment.
- Existing studies on differential privacy assume a simple data model (statistics single or multidimensional numerical spaces). Mechanisms to extend differential privacy to more complicated data domains (graphs or strings) or complex query plans (recursive SQL queries) need to be explored.

V. Domain Level Risks And Remedies

A. Web search engines

Online search engines are tools that help users find information. These search engines use the information provided by users, in terms of their search history to build their "user profiles". Rich user profiles enable the search engines to provide better personalized search results.

However, this puts the user's privacy at risk. Apart from the risk of exposing one's identity, there is the added disadvantage of being subjected to unsolicited advertising and unwanted disclosure of sensitive information. Rich user profiles contain a lot of personally identifiable information, which can attract unwarranted malicious interests.

A search engine is only as successful as the number of useful search results it provides. Search engines and recommendation systems benefit from users who make personal information available, thereby providing tailor-made search results and/or recommendations. However a user risks his/her privacy to gain personalized recommendations.

Sanchez et al. in [33], propose a mechanism where users are in control of the amount of private information they reveal vs. the degree of richness their user profiles retain. Their solution obfuscates the original query by adding "k" number of fake queries to the original query and then submitting the entire set to the Web Search Engines (WSE). However unlike solutions like TrackMeNot which generate "k" random queries, their solution generates "k" queries that are semantically related to the original query. The semantic distance ensures that the fake queries do not deviate far away from the interests of the user thereby ensuring the user profiles are still meaningful. Users control the number of fake queries added as well as the amount of semantic distance between the original and fake queries.

The contributions of Sanchez et al. include the proposal of a new scheme to generate distorted queries from a semantic point of view, thereby preserving utility of user profiles. Their work also supported complex queries and provided a tradeoff between Utility and Privacy addressed via configurable parameters.

While the contributions of their work towards providing users with control over their privacy preservation and personalization are exciting, the performance impact is not studied in detail. The solution makes use of knowledge bases like WordNet and Open Data Project (ODP) to extract query topics and concepts that are at a given semantic distance from the query topic. This is an important step in their obfuscation technique. While the average time taken for a query found in WordNet is 30 ms, it is 1500 ms for one not found in WordNet and found in ODP in the second iteration. Also, the approach obfuscates the search queries, but the submission of these queries to search engine is not linked to a particular protocol.

Viejo et al. in [42] propose profiling users locally based on their social network account usage instead of their web search history. They then use the local profile as a base to obfuscate the user's search queries. By not using a user's web search history to create a user profile, this approach forces search engines to focus on a user's "macro" interests instead of their fine-grained "micro" interests.

However local profiles built thus are static and may not represent a user's true interests over time. Also, query obfuscation in this approach, is based on a high-level characterization of user's interests and ignores the semantic preservation of the original query.

Hassan et al. in [12] discuss the ongoing European Union funded EEXCESS[2] project as an example of providing improved user recommendations by making use of intensive user profiling techniques. One of the major challenges is that the EEXCESS architecture is based on a federated recommender system in which future partners may join. The trustworthiness and the intent of these partners are not necessarily known. The information collected and disclosed to recommenders may not, in itself, be sensitive; however, cross-referencing it with external big data sources and analyzing it through big data techniques may create breaches in user privacy. Since, untrustworthy partners may have access to such big data sources and analysis techniques that privacy becomes a clear challenge. The EEXCESS project addresses the challenges of guaranteeing privacy, based on flexible privacy policies and evaluating the trust and reputation of a recommender.

Although EEXCESS project proposes a novel user-controlled approach to privacy preserving searches that ensure rich recommendations as results, it requires a complete architectural change to how web searches are done.

B. Social Networking

In the context of this paper we refer to sites such as Facebook, Linked in, Flicker etc., as social networks. These are sites where users can befriend other entities and be linked to and interact with them, tag them, discuss and present opinions/preferences/reviews.

As of June 2014, on an average 829 million users were active on Facebook per day. Monthly active user counts go up to 1.07 billion [29]. Social networking sites have proliferated into our lives like never before. While these sites bring indisputable convenience of staying connected, they often bring with them several privacy challenges. Disclosing information on such platforms is usually with consent, but users are often unaware of privacy settings and are not sure if and how their data will be used.

As noted by Zheleva and Getoor in [47] privacy in social networks can be discussed under two scenarios: privacy breaches and data anonymization.

1. Privacy breach: To quote Zheleva et al., "Privacy breach occurs when a piece of sensitive information about an individual is disclosed to an adversary, someone whose goal is to compromise privacy." Privacy breach includes the following type of disclosures:

- Identity disclosure: Identity disclosure happens when an adversary is able to map a social network profile to a real world entity.
- Attribute disclosure: Attribute disclosure occurs when an adversary is able to determine the sensitive attribute value. A sensitive attribute is one which the user intended to keep private.
- Social link disclosure: Social link disclosure happens when an adversary learns about a sensitive link between two users (which the users intended to keep private).

Affiliation link disclosure: Affiliation link disclosure happens when an adversary is able to determine if a user belongs to a particular affiliation group. Affiliation link disclosures often lead to attribute, social link and identity

[2] eexcess.eu

disclosures. Therefore it is important to preserve the privacy of a user's affiliations.

2. Data Anonymization: This presents a scenario when data provider would want to release data to researchers, keeping in mind user's privacy. Privacy preservation often comes down to how well one can hide in a crowd. In order to accomplish this, the data provider would have to perturb the data by generalization and/or anonymization. Anonymizing social network data presents more challenges than anonymizing tabular data (used in traditional data publishing). This is because in the social network context one has to anonymize the attributes as well as the network structure. Some of the methods discussed by Zheleva et al. in [47] are listed below:

- *Naive Anonymization of network structure:* A naive way of anonymizing a social network structure is to empty the profile of all of its attributes and leave only the link structure. However such a structure is still susceptible to attacks where an adversary remembers the structure of his own network and has created a particular pattern of links to the nodes of interest before the structure is anonymized.
- *K-degree anonymity:* Having background knowledge about the sub-graph structure of "a node of interest" enables an adversary to identify that node in an anonymized network. Here the node's links (or degree) become tools that reveal its identity and should therefore be hidden. *k*-degree anonymity states that each node should have at least *k*-1 other nodes with the same degree, in an anonymized network.
- *K-neighborhood anonymity:* An adversary having background information about a node's 1.5 hop neighborhood can easily identify the node in an anonymized network. To overcome such attacks *k*-neighborhood anonymity is used. *k*-neighborhood anonymity states that each node should have a 1.5 hop neighborhood graph that is isomorphic to the 1.5 hop neighborhood graphs of at least *k*-1 other nodes in the anonymized network.
- *K-candidate anonymity:* This is a more general privacy preservation measure which takes care of background attacks regardless of the size of the neighborhood information possessed by the adversary. It states that an anonymized graph satisfies *k*-candidate anonymity if for any structural query Q posed to this network, there is at least a set of *k* nodes that match Q.
- *Anonymization of attributes:* During attribute anonymization each node in the social network is treated as a record having several attributes and these attributes are anonymized by way of generalization and suppression. Some of the techniques used are *k*-anonymity, *l*-diversity, *t*-closeness and differential privacy.

Privacy preservation in social networks is relatively new in the realm of data privacy research. Research has begun to understand and overcome some of the challenges. However more work towards comparing anonymization techniques in terms of utility preservation remains to be done. Also exciting is the study of user-controlled privacy vs. utility preservation such that individual requirements can be met.

C. Healthcare

A lot of valuable data is generated between patient visit records, prescription history, immunization and lab records and interactions with health care providers. Storing such data in-house is sometimes not an option for many healthcare providers. The trend is to shift towards third party storage and management of Personal Health Records (PHR), where service providers like Microsoft Health Vault store and manage PHR data interactions. PHR services like these are convenient because they enable efficient storage and sharing of health records.

However, there are several security and privacy risks associated with these solutions which unless mitigated make these solutions less attractive. PHRs contain Personally Identifiable Information (PII) and many healthcare providers as well as patients are concerned about this being stored in third-party sites. These sites also become targets of malicious attacks because of the sensitive information they store.

Samavi et al. in [32], an empirical study on "PHR User Privacy Concerns and Behaviors", found out that in-spite of having privacy concerns, 60% of the study-respondents, reported not reading privacy agreements. PHR users who were highly concerned about privacy did not change their default privacy settings. Therefore the authors conclude that PHR architects and developers cannot rely on "privacy agreements" alone and identify a critical need for adopting alternative privacy preserving options.

Anonymizing and encrypting the files before outsourcing them seems like an appropriate and feasible solution. Giving the patients the control of how to encrypt their files and whom to give access to what part of their PHR is a novel approach. Heurix et al. in [13] discuss the prejudice and discrimination patients are subjected to in the event of unauthorized disclosure of their Electronic Health Records (EHR) to insurance companies or even potential employers. The authors present a pseudonymization approach that preserves the patient's privacy and data confidentiality and provides a balance between privacy and data usability. Pseudonymization combines the strength of anonymization and full document encryption. It separates medical content (like x-ray images) from personally identifiable information (PII) like name, date of birth etc. Both records are assigned randomly selected pseudonyms. These pseudonyms act as access tokens: knowing the correct pseudonym will help re-link the health records with the corresponding patient. The pseudonyms are protected with encryption using a user-specific secret key. Unlike anonymization, pseudonymization is reversible only by authorized users who possess the secret key. Pseudonymization relies on encryption, but unlike full document encryption, here only the metadata is encrypted, considerably reducing the cryptography overhead.

Ming Li et al. [19] propose a patient centric Attribute Based Encryption (ABE) framework, for sharing PHR in cloud environments, under multi-owner settings. The framework lets patients the right to grant and revoke access to their PHR files. In order to address scalable key management, the authorized users are divided into two domains - Public (hospitals, doctors, pharmacists, researchers) and Personal (friends and family).

The public domain will have multiple Attribute Authorities (AAs) like Hospitals, Pharmacies, Insurance Companies etc. Readers from each domain get their secret keys (role based attributes), from their respective domain AAs. For example, a doctor in the hospital domain will obtain his secret key from the Hospital AA. This secret key is based on the doctor's role, and may look like "hospital H1, surgeon, M.D., cardiology". Here Multi Authority ABE (MA-ABE) is used to encrypt data. In the personal domain, PHR owner defines a set of data attributes like "personal information", "medical history", "allergies", "prescription" etc. A user in the personal domain can send a message to the PHR owner and request access to a set of attributes.

The PHR owner then grants the requester access to a subset of data attributes. Here Key Policy ABE (KP-ABE) is used to encrypt data. PHR owners' files are encrypted with both fine-grained (personal domain) and role-based (public domain) access policy. Once encrypted only authorized users with the required attribute based access can decrypt these files.

While this is a novel approach to give user's control of who they share their health data with, it is not clear how well this fares when the number of PHR users grows in the public and the personal domain. It is yet to be seen if the solution scales efficiently.

D. Context Aware Mobility

As per the 2014 Global Internet Report released by the Internet Society [34], number of internet users over mobile broadband exceeded fixed users four years ago in February 2010. With the advent of smart phones, accessing the internet over mobile devices has seen an unprecedented surge in the recent years. Mobile devices are capable of providing rich location based information because they are equipped with several sensors like camera, GPS, microphone, Wi-Fi, accelerometers, gyroscopes etc. These sensors are capable of capturing user context information such as location, temperature, motion and other entities in the vicinity. Location based and context based personalization services use this context (mainly location and time information) to provide tailor made recommendations. While this is a cool utility, it does put the mobile user's privacy at risk.

Most of the work done in privacy preservation of context aware applications focuses on location and activity inference. As indicated by Jagtap et al. in [17], the existing controls on context-aware systems are static and predetermined. On most smart phones a user is asked permission to share sensor information such as location at install time. This is not enough because context is dynamic and in itself can determine what data should be shared. In fact users should be given the control of deciding what sensor information they would like to share, with whom and under what context. Jagtap et al [17], capture these requirements in their on-going project where they built a policy based framework that uses semantic reasoning to control the flow of information from the sensors. It uses OWL ontology to represent dynamic context aware system and a combination of OWL-DL and Jena rules to specify privacy policies. The system uses automatic generalization of context attributes (like building name instead of the exact GPS coordinates) based on user specified privacy policy.

Web Ontology Language (OWL) [25], a W3C standard, is a language for processing web information. It is written in XML and is designed to be interpreted by computers. OWL-DL [25] is a sub language of OWL and supports those users who want the maximum expressiveness without losing computational completeness (all entailments are guaranteed to be computed) and decidability (all computations will finish in finite time) of reasoning systems. OWL DL was designed to support the existing Description Logic business segment and has desirable computational properties for reasoning systems. Jena [24] is an API in the Java programming language, for the creation and manipulation of RDF graphs. Jena was developed to satisfy two goals: to provide an API that was easier for the programmer to use than alternative implementations and to be conformant to the RDF specifications. Resource Description Framework (RDF) is a W3C standard for describing web content. RDF describes a web resource with its title, author, copyright information and its contents.

Cornelius et al. in [5] proposed AnonySense, a privacy aware architecture, in which sensor data is anonymized before sharing it. Context Privacy Services (CoPS) [31] puts user in control of when, how and with whom a user's context will be shared but it does not cater to context-dependent privacy policies.

While solutions in context-aware privacy preservation aim to achieve user controlled context sharing, the scalability, performance and utility of these solutions needs to be established and is the focus of future research.

E. Wearable Technology

Wearable technology or wearable devices refer to small electronic computers that are incorporated into clothing or accessories and are worn on the body [18]. Wearable technology is often more sophisticated than hand held computers in terms of bio-feedback and real-time communication capability. Examples of such technology includes and is not limited to glasses, watches, on-body life logging cameras, hearing aids, heart-rate monitors, vital sign recorders and smart fabric. Some of the more invasive wearable technology includes devices that are implanted inside a human body. As discussed by Davenport in [6], wearable and portable Haemodialysis devices are now under clinical trial. With advancement in nanotechnology, electronics and miniaturization, research is also underway to develop implantable Haemodialysis devices which may revolutionize the treatment and quality of life for patients with end-stage kidney disease. Add to this the research at Google X to manufacture an implantable technology for early cancer detection [43], where users will ingest a pill that releases nano particles and attaches itself to body cells and proteins, which subsequently reports health information to a wearable device. With this research Google is trying to change medicine from reactive and transactional to proactive and preventative.

Life loggers include on-body passive cameras and microphones, which could essentially be "always on" and have the ability to act as long-term archiving devices and memory aids. Some examples of logging devices are the Narrative Clip[3], Autographer[4], Google Glass [35] and Microsoft's SenseCam [30], which are now affordable and available for public use. The Narrative Clip is a small wearable camera that includes location sensing. It is 1.42 x 1.42 x .35 inches and weighs approximately 0.7 ounces. The battery life of one charge of the Narrative Clip is around 24-30 hours, enough for two days of use before you need to recharge the Clip. This is for the default setting of one photo every 30 seconds. The clip costs anywhere from $229 to $279 and can be bought off getnarrative.com [44]. The Autographer is slightly larger than the narrative clip in form and has an in-built GPS sensor. It costs about $399 and can be bought via autographer.com [44]. Google Glasses look like a pair of eyeglasses, but the lenses of the glasses are an interactive, smart phone-like display, with natural language voice command support as well as Blue-tooth and Wi-Fi connectivity. Google Glass is powered by the Android mobile operating system and compatibility with both Android-powered mobile devices and Apple iOS-powered devices is expected. One can become an explorer and buy glass by visiting one of the "Glass Base Camps" in the United States. Glass was sold for about $1500 in 2013[35]. Microsoft's SenseCam (now available to buy as Vicon Revue), has been extensively used in retrospective memory assistance. SenseCam is a wearable camera that takes photos automatically [30].

As noted by Wolf et al. in [44], the reasons for using such devices are manifold: to provide healthcare assistance (in the form of personal memory aids or introspection), or simply to record memorable events (such as a holiday trip or a birthday). Hodges et al. in [14] discuss the use of Microsoft's SenseCam as a retrospective memory aid in a 12-month clinical trial with a patient suffering from Amnesia. Their results indicate periodic review of images captured by SenseCam significantly improved the recall rate by the patient which was previously impossible. Vallurupalli et al. in [41] discussed the feasibility of using Google Glass to explore different scenarios in cardiovascular fellowship, where fellows can better their education by getting real-time and appropriate supervision by experienced faculty. The mock trainee wore a Google Glass and the live video stream from Glass was fed via Blue tooth or Wi-Fi to a smart device for the supervisor to view and advice. Through this study it was concluded that wearable technology can enhance medical education once available widely.

While one cannot dispute the positive impact wearable technology has had in the medical domain, issues related to privacy continue to surface, especially with respect to who has access to the data collected by these devices and how it is communicated among various data recipients. Wearable camera devices also capture tons of images, many of which may put the participant's as well as a bystander's privacy at risk. To address ethical and privacy issues with the use of wearable cameras in health care and research, Kelly et al. in [3], attempt to formalize protection for all under best ethical practices by developing an ethical framework.

However, majority of privacy concerns with wearable technology, are in its non medical use, especially with the "life logging" function and the long term archiving facility that it brings with it. Intille et al. in [16] depicted several scenarios where the use of wearable technology can intrude the wearer's as well as a bystander's privacy and discussed how social expectations as well as privacy laws need to change, to prevent privacy erosion, as the use of wearable technology becomes pervasive in our society.

Significant research is underway to address privacy issues with wearable devices. Yus et al. developed FaceBlock [46], as a solution to preserve privacy of users in the scenario where eye wear devices are raising privacy concerns among the general public. FaceBlock allows users to state their policy about being photographed (allow vs. disallow pictures) by other people. FaceBlock generates an Eigen face, a mathematical representation or a face identifier, using a picture of a user's face. Whenever a Glass user is in the vicinity of the user, FaceBlock forms an adhoc connection with it and sends the face identifier along with the policy. In order to enforce the policy, the FaceBlock application running on Google Glass uses the face identifier to detect if the user who shared the policy is part of the pictures taken by the device. It then selectively obscures the face of all the people who have sent such a policy to the device.

Pappachan et al. in [28], describe a context-aware privacy model (represented using OWL[25] ontologies and SWRL[15] rules), that helps FaceBlock to generate Privacy-Aware pictures depending on the context and privacy needs of a user instead of the all-or-nothing model of the original FaceBlock. Web Ontology Language (OWL), a W3C standard, is a language for processing web information. It is written in XML and is designed to be interpreted by computers. Semantic Web Rule Language (SWRL) includes a high-level abstract syntax for Horn-like rules and logic. Pappachan et al. in [28] model a user's context (identity, activity, location and time), in an ontology and encode a users privacy policy as a SWRL rule.

Templeman et al. introduce PlaceAvoider [38], a technique for owners of a wearable camera to blacklist sensitive places (bathrooms, bedrooms etc.) that they want to exclude from their life logging images. With PlaceAvoider, wearers submit pictures of sensitive places

[3] getnarrative.com

[4] autographer.com

and the system builds visual models of rooms that should not be captured. PlaceAvoider later on recognizes the images taken in these areas and flags them for further user review helping the owner to either discard or blur such images.

Use of First Person Point of View (FPPOV) imagery (via wearable cameras) in behavioral research is on the rise. Thomaz et al. in [40], address the privacy concerns of FPPOV imagery by providing a formulation around such images to classify them as "images that might pose a privacy concern" vs. "images which contain salient information with respect to a particular behavioral research task". The formulation centers on a 2 x 2, Privacy-Saliency matrix, which helps take next steps of either discarding images with privacy concern altogether or retaining the images after blurring the areas with privacy concern. The authors also combine sensor data such as geo locations and data streams collected from mobile phone accelerometers to assist location and motion filtering techniques to address privacy concerns in FPPOV imagery.

F. Internet of Things

With the advent of Radio Frequency Identification (RFID) technology, which enables objects with unique identifiers to transmit data and communicate with other objects on a real-time basis, the internet of things has become a reality. As noted by Charu et al. in [1], the number of interconnected devices outran the number of humans on this planet as of 2008. Real-time communication between these interconnected devices generates enormous amount of data whose collection and processing is a challenge. This challenge is surmounted by the use of big data analytics. However, such data presents several privacy risks because it often contains identifiable and sensitive information. As indicated by Charu et al. in [1], there are privacy risks during data collection, data transmission and data mining.

RFIDs can be tracked and if these RFIDs are on a person, the Electronic Product Code (EPC) emitted by the device becomes a unique identifier of that person. This data collection process poses risk of exposing the whereabouts of the individual. Charu et al. in [1] propose several solutions to mitigate this risk.

Protect RFID privacy by way of the kill command. The password protected kill command can be triggered by a signal at point of sale. It disables the tag and the device no longer emits EPC. Although this works for products with short life span, it may not work with smart products, which need to function throughout their life time.

Have tags that have locking and unlocking mechanism. The tags will emit EPC only during specified data collection times, when the environment is known to be secure. Security during data transmission can be handled by encrypting the EPC before transmission. However this is limited in application because the encryption only protects the contents of the tag and not the tag itself.

Embed dynamic encryption ability within the tag. The authors cite some of the cryptographic schemes with re-writable memory inside the tags. Readers can encrypt a tag and write it back to this memory (every time a tag is read and decrypted), so as to confuse an eavesdropper with different encrypted tags at different times. This solution comes at considerable cryptography cost.

Use of blocker tags which spam unauthorized readers so as to cause such readers to stall.

Mask the RFID tags with a set of pseudonyms and transmit the pseudonym instead of the EPC. This makes is difficult for malicious parties to identify the tags because the tags are associated with different pseudonyms at different times.

Charu et al. in [1], also discuss mechanisms to maintain privacy during data sharing. These techniques include aggregation and reducing the accuracy of the data by *k*-anonymity, *l*-diversity, *t*-closeness and differential privacy. A novel approach would be to provide users control over what information is shared, with whom and under what context. This idea has also been addressed by Charu et al. in [1], in the context of semantic web, where apart from specifying access control, a user can control fine-tuned granularity of query responses depending upon the identity of the entity requesting the query and its context.

Location is a context-rich piece of information that is both sensitive and necessary for many IoT scenarios such as participatory sensing or location-based online services. Agir et al. in [2], discuss a general approach to preserve location privacy which is commonly tracked by location sensing devices and applications that feed the Internet of Things (IoT). They propose an adaptive location privacy protection scheme that takes into account not only the geographical but also the semantic information of urban locations, as well as user's sensitivities to obfuscate the location information that is transmitted to a service provider of Internet of Things. Their scheme also emulates a sophisticated adversary attack and thereby estimates the users expected privacy levels. In essence, the protection mechanism iteratively adjusts its obfuscation parameters until the user's sensitivity preferences are satisfied.

Gudymenko et al. in [11] suggest that mere technological solutions will not be sufficient to mitigate the privacy risks associated with the Internet of Things (IoT). They opine that privacy regulation and legal enforcement of privacy rights is vital for the management of privacy in the realm of IoT. The authors indicate that the following requirements must be considered when approaching privacy risks associated with RFID based IoT:

- *Assessment of privacy compliance of the RFID system:* via Privacy Impact Assessment (PIA) certification.
- *Ability to opt-out:* Users should be able disable communication of their RFID devices at any time.
- *Ability to permanently disable the tag:* Some tags can be remotely reactivated even after being temporarily disabled using the "kill command". Therefore mechanism should be provided to permanently disable or physically destroy the tag.
- *Marking the intelligence-enabled artefacts:* Attaching special markers to smart things to indicate their IoT activity.

- *Considering M2M Privacy:* Machine to Machine privacy and not just individual-related privacy must to be considered.
- *RFID usage restriction:* Sensitive areas (AIDS centres for example) should have prohibited or restricted use of RFID technology.

VI. Conclusion And Future Work

This survey presents the opportunities and challenges that are associated with being "connected" in the digital world. This paper focuses on bringing out various privacy challenges that are present in all areas of internet usage including Web Search, Social Networking, Healthcare, Mobility, Wearable Technology and Internet of Things. Several widely discussed and adopted privacy preservation mechanisms are presented for each domain. Limitations of existing solutions and the scope of future work are also discussed.

One problem that is gaining continued focus is privacy preserving web search and selective personalization where the user is in control of the degree of personalization vs. privacy. We intend to tackle this problem in our future work where a user can conduct anonymous web searches by way of crowd hiding. This will be achieved by obfuscating original search queries with semantically related fake queries. Users will specify the amount of obfuscation that can be applied to their search queries in terms of the number of fake queries added and the semantic distance between the original and fake queries. Users will also be able to decide what part of their user profile they are willing to share with recommenders. They will be able to play with these settings and be provided with instant feedback on how their privacy settings affect the quality of their search results so they can make an informed decision.

References

[1] Charu C. Aggarwal, Naveen Ashish, and Amit Sheth. The internet of things: A survey from the data-centric perspective. In Charu C. Aggarwal, editor, Managing and Mining Sensor Data, pages 383–428. Springer US, 2013.

[2] Berker Agir, Jean-Paul Calbimonte, and Karl Aberer. Semantic and Sensitivity Aware Location-Privacy Protection for the Internet of Things. In Privacy Online: Workshop on Society, Privacy and the Semantic Web Privon 2014, 2014.

[3] Kelly Paul Marshall Simon J. Badland Hannah Kerr Jacqueline Oliver Melody Doherty Aiden R. Foster Charlie. An ethical framework for automated, wearable cameras in health behavior research. Elsevier, 44, November 2014.

[4] Chris Clifton. Privacy-preserving data mining. In LING LIU and M.TAMER ÃU˝ ZSU, editors, Encyclopedia of Database Systems, pages 2147–2150. Springer US, 2009.

[5] Cory Cornelius, Apu Kapadia, David Kotz, Dan Peebles, Minho Shin, and Nikos Triandopoulos. Anonysense: Privacy-aware people-centric sensing. In Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services, MobiSys '08, pages 211–224, New York, NY, USA, 2008. ACM.

[6] Andrew Davenport. Portable and wearable dialysis devices for the treatment of patients with end-stage kidney failure: Wishful thinking or just over the horizon? Pediatric Nephrology, pages 1–8, 2014.

[7] Josep Domingo-Ferrer and Vicenç Torra. A critique of k-anonymity and some of its enhancements. In Proceedings of the 2008 Third International Conference on Availability, Reliability and Security, ARES '08, pages 990–993, Washington, DC, USA, 2008. IEEE Computer Society.

[8] Charles Duhigg. How companies learn your secrets. N.Y. TIMES MAGAZINE, Feb 2012.

[9] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, Automata, Languages and Programming, volume 4052 of Lecture Notes in Computer Science, pages 1–12. Springer Berlin Heidelberg, 2006.

[10] José Luis Fernández-Alemán, Inmaculada Carrión Señor, Pedro Ángel Oliver Lozoya, and Ambrosio Toval. Security and privacy in electronic health records: A systematic literature review. Journal of biomedical informatics, 46(3):541–562, 2013.

[11] Ivan Gudymenko, Katrin Borcea-Pfitzmann, and Katja Tietze. Privacy implications of the internet of things. In Constructing Ambient Intelligence, pages 280–286. Springer, 2012.

[12] Omar Hasan, Benjamin Habegger, Lionel Brunie, Nadia Bennani, and Ernesto Damiani. A discussion of privacy challenges in user profiling with big data techniques: The eexcess use case. In Proceedings of the 2013 IEEE International Congress on Big Data, BIGDATACONGRESS '13, pages 25–30, Washington, DC, USA, 2013. IEEE Computer Society.

[13] Johannes Heurix and Thomas Neubauer. Privacy-preserving storage and access of medical data through pseudonymization and encryption. In Trust, Privacy and Security in Digital Business, pages 186–197. Springer, 2011. 10

[14] Steve Hodges, Lyndsay Williams, Emma Berry, Shahram Izadi, James Srinivasan, Alex Butler, Gavin Smyth, Narinder Kapur, and Ken Wood. Sensecam: A retrospective memory aid. In Paul Dourish and Adrian Friday, editors, UbiComp 2006: Ubiquitous Computing, volume 4206 of Lecture Notes in Computer Science, pages 177–193. Springer Berlin Heidelberg, 2006.

[15] Ian Horrocks, Peter F Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, et al. Swrl: A semantic web rule language combining owl and ruleml. W3C Member submission, 21:79, 2004.

[16] A.M. Intille and S. S. Intille. New challenges for privacy law: Wearable computers that create electronic digital diaries. Massachusetts Institute of Technology, Cambridge, MA, MIT Dept. of Architecture House Project Technical Report, 2003.

[17] P. Jagtap, A. Joshi, T. Finin, and L. Zavala. Preserving privacy in context-aware systems. In Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on, pages 149–153, Sept 2011.

[18] Tehrani Kiana and Andrew Michael. Wearable technology and wearable devices: Everything you need to know. Wearable Devices Magazine, WearableDevices.com, March 2014.

[19] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. Parallel and Distributed Systems, IEEE Transactions on, 24(1):131–143, Jan 2013.

[20] Ninghui Li, Tiancheng Li, and S. Venkatasubramanian. T-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on, pages 106–115, April 2007.

[21] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond kanonymity. ACM Trans. Knowl. Discov. Data, 1(1), March 2007.

[22] Mary Madden. Public perceptions of privacy and security in the post snowden era. Pew Research Center, November 2014.

[23] Paul Martini. A secure approach to wearable technology. Network Security, 2014(10):15–17, 2014.

[24] Brian McBride. Jena: Implementing the rdf model and syntax specification. In SemWeb, 2001.

[25] Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. W3C recommendation, 10(10):2004, 2004.

[26] Jakub Mikians, L ászl ó Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. Detecting price and search discrimination on the internet. In Proceedings of the 11th ACM Workshop on Hot Topics in Networks, HotNets-XI, pages 79–84, New York, NY, USA, 2012. ACM.

[27] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In Proceedings of the 2008 IEEE Symposium on Security and Privacy, SP ’08, pages 111–125, Washington, DC, USA, 2008. IEEE Computer Society.

[28] Primal Pappachan, Roberto Yus, Prajit Kumar Das, Tim Finin, Eduardo Mena, and Anupam Joshi. A Semantic Context-Aware Privacy Model for FaceBlock. In Second International Workshop on Society, Privacy and the Semantic Web - Policy and Technology (PrivOn 2014), Riva del Garda (Italy), October 2014.

[29] Facebook Investor Relations. Facebook reports second quarter 2014 results. Facebook.

[30] Microsoft Research. Sensecam. Web.

[31] Vagner Sacramento, Markus Endler, and Fernando Ney Nascimento. A privacy service for context-aware mobile computing. In Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks, SECURECOMM ’05, pages 182– 193, Washington, DC, USA, 2005. IEEE Computer Society.

[32] Reza Samavi, Mariano P. Consens, and Mark Chignell. {PHR} user privacy concerns and behaviours. Procedia Computer Science, 37(0):517 – 524, 2014. The 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014)/ The 4th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2014)/ Affiliated Workshops.

[33] David S áNchez, Jordi Castell í-Roca, and Alexandre Viejo. Knowledge based scheme to create privacy-preserving but semantically-related queries for web search engines. Inf. Sci., 218:17–30, January 2013.

[34] Internet Society. Global internet report 2014. Web, 2014.

[35] Forrest Stroud. Google glass. Web.

[36] LATANYA SWEENEY. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05):557–570, 2002.

[37] C. Task and C. Clifton. A guide to differential privacy theory in social network analysis. In Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on, pages 411– 417, Aug 2012.

[38] Robert Templeman, Mohammed Korayem, David Crandall, and Apu Kapadia. Placeavoider: Steering first-person cameras away from sensitive spaces. In Network and Distributed System Security Symposium (NDSS), 2014.

[39] Omer Tene and Jules Polonetsky. Big data for all: Privacy and user control in the age of analytics.

[40] Edison Thomaz, Aman Parnami, Jonathan Bidwell, Irfan Essa, and Gregory D. Abowd. Technological approaches for addressing privacy concerns when recognizing eating behaviors with wearable cameras. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp ’13, pages 739–748, New York, NY, USA, 2013. ACM.

[41] S. Vallurupalli, H. Paydak, S.K. Agarwal, M. Agrawal, and C. Assad- Kottner. Wearable technology to improve education and patient outcomes in a cardiology fellowship program - a feasibility study. Health and Technology, 3(4):267–270, 2013.

[42] A. Viejo and D. Sanchez. Providing useful and private web search by means of social network profiling. In Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on, pages 358–361, July 2013.

[43] Alistair Barr & Ron Winslow. Google’s newest search: Cancer cells. The Wall Street Journal, October 2014.

[44] K. Wolf, A. Schmidt, A. Bexheti, and M. Langheinrich. Lifelogging: You’re wearing a camera? Pervasive Computing, IEEE, 13(3):8–12, July 2014.

[45] Yin Yang, Zhenjie Zhang, Gerome Miklau, Marianne Winslett, and Xiaokui Xiao. Differential privacy in data publication and analysis. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD ’12, pages 601–606, New York, NY, USA, 2012. ACM.

[46] Roberto Yus, Primal Pappachan, Prajit Kumar Das, Eduardo Mena, Anupam Joshi, and Tim Finin. FaceBlock: Privacy-Aware Pictures for Google Glass. In Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys ’14, page 1. ACM SIGMOBILE, June 2014.

[47] Elena Zheleva and Lise Getoor. Privacy in social networks: A survey. In Charu C. Aggarwal, editor, Social Network Data Analytics, pages 277–306. Springer US, 2011

[48] Weber, Rolf H. "Internet of Things–New security and privacy challenges."Computer Law & Security Review 26.1 (2010): 23-30.

[49] Zhang, Chi, et al. "Privacy and security for online social networks: challenges and opportunities." *Network, IEEE* 24.4 (2010): 13-18.

[50] Fung, Benjamin, et al. "Privacy-preserving data publishing: A survey of recent developments." *ACM Computing Surveys (CSUR)* 42.4 (2010): 14.

[51] Toch, Eran, Yang Wang, and Lorrie Faith Cranor. "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems." *User Modeling and User-Adapted Interaction* 22.1-2 (2012): 203-220.

Author Profiles

Arti Arya has completed BSc (Mathematics Hons) in 1994 and MSc (Mathematics) in 1996 from Delhi University. She has completed her Doctorate of Philosophy in Computer Science Engineering from Faculty of Technology and Engineering from Maharishi Dayanand University, Rohtak, Haryana in 2009. She is working as Professor and Head of MCA dept in PESIT, Bangalore South Campus. She has 15 yrs of experience in academics, of which 7 yrs is of research. Her areas of interest include spatial data mining, knowledge based systems, text mining, unstructured data management, knowledge based systems, machine learning, artificial intelligence, applied numerical methods and biostatistics. She is a life member of CSI and member IEEE. She is on the reviewer board of many reputed International Journals.

Saraswathi Punagin is a graduate student at the Department of Computer Science and Engineering in PESIT, Bangalore South Campus. She has over 15 years of experience in the field of Information Technology and has worked in diverse industries such as healthcare, government, telecommunications and technical consulting. During her tenure she has worked in capacities of solutions developer, technical writer, data and business intelligence analyst, project manager as well as subject matter expert for various business and operational intelligence solutions. Her areas of interest include data privacy, big data analytics and data science.