Noname manuscript No.
(will be inserted by the editor)

A Bibliometric Analysis of Privacy and Ethics in IEEE Security & Privacy

Jonathan Tse · Dawn E. Schrader · Dipayan Ghosh · Tony Liao · David Lundie

the date of receipt and acceptance should be inserted later

Abstract The increasingly ubiquitous use of technology has led to the concomitant rise of intensified data collection and the ethical issues associated with the privacy and security of that data. In order to address the question of how these ethical concerns are discussed in the literature surrounding the subject, we examined articles published in *IEEE Security & Privacy*, a magazine targeted towards a general, technically-oriented readership spanning both academia and industry.

Our investigation of the intersection between the ethical and technological dimensions of privacy and security is structured as a bibliometric analysis. Our dataset covers all articles published in *IEEE Security & Privacy* since its inception in 2003 to February 06, 2014. This venue was chosen not only because of its target readership, but also because a preliminary search of keywords related to ethics, privacy, and security topics in the ISI Web of Knowledge and IEEE Xplore indicated that *IEEE Security & Privacy* has published a preponderance of articles matching those topics. In fact, our search returned two-fold more articles for *IEEE Security & Privacy* than the next most prolific venue. These reasons, coupled with the fact that both academia and industry are well-represented in the authorship of articles makes *IEEE Security & Privacy* an excellent candidate for bibliometric analysis.

J. Tse · D.E. Schrader · D. Ghosh
Cornell University
Ithaca, NY,USA
E-mail: {jdt76,des14,dpg65}@cornell.edu

T. Liao
Media Studies and Production, Temple University
Philadelphia, PA 19147, USA
E-mail: tony.liao@temple.edu

D. Lundie
Liverpool Hope University
Hope Park, Liverpool, UK
E-mail: lundied@hope.ac.uk

Our analysis examines the ways articles in *IEEE Security & Privacy* relate ethics to information technology. Such articles can influence the development of law, policy and the future of information technology ethics. We employed thematic and JK-biplot analyses of content relating privacy and ethics and found eight dominant themes as well as the inter-theme relationships. Authors and institutional affiliations were examined to discern whether centers of research activity and/or authors dominated the overall field or thematic areas. Results suggest avenues for future work in critical areas, especially for closing present gaps in the coverage of ethics and information technology privacy and security themes particularly in the areas of ethics and privacy awareness.

Keywords Ethics · Privacy · Education

1 Introduction

The collection of personal information and the technological means for processing and leveraging it has expanded rapidly over the past decade (Pham et al, 2011; Richards and King, 2014). Whether through web browsing, locative surveillance devices such as cell phones and CCTV, or in-home utility monitoring technologies, people are disclosing personal information. Big Data and profiling, especially for economic gains (Diebold, 2003), have effectively increased the amount of this personal data collected and processed. Individual awareness and decision making around privacy and security issues have grown more important as well. While both industry and individuals benefit from big data, any benefits or conveniences may ultimately compromise the security and privacy of individuals (Acquisti, 2014). In fact, there are far reaching effects that extend beyond economic considerations such as negative influences on decision making, autonomy, and self regulation (Cohen, 2013; Wicker and Schrader, 2011)

These effects have created challenges for our legal systems and design principles (Lessig, 2006; Toubiana et al, 2011), technological designers (Gürses et al, 2011; Hull et al, 2011; Shilton, 2010), individual reputations (Solove, 2007), decision-making processes (Acquisti and Grossklags, 2005; Sar and Al-Saggaf, 2014), and for society in terms of surveillance and dataveillance (Turow, 2012). Balancing the competing concerns of privacy, convenience, and security is of critical importance at both the societal and individual level. With the rise of data mining practices, the ethics surrounding data sharing and aggregation on security and privacy are of critical importance.

Awareness of privacy concerns is increasing among technology users, but mostly as a reaction to high-profile data confidentiality breaches. In contrast, information systems designers and analysts must take a proactive role with respect to privacy, especially with regards to the ethical implications of their work and its effects on consumers. The understanding of ethical issues related to privacy is still in its infancy. This is partly due to the changing scope and definition of what privacy is, including what it means legally and socially. One working definition has been put forth by legal scholar Julie Cohen:

> Privacy is shorthand for breathing room to engage in the processes of boundary management that enable and constitute self-development. So understood, privacy is fundamentally dynamic. In a world characterized by pervasive social shaping of subjectivity, privacy fosters (partial) self-determination. It enables individuals both to maintain relational ties and to develop critical perspectives on the world around them. (Cohen, 2013)

A key problem is that people are often unaware of the value of their private information. To further complicate matters, the nature of what is considered private is changing as social and technological worlds change (Hugl, 2010; Toubiana et al, 2010). There has been a quantitative and qualitative shift in how new information and communication technologies have complicated the study of privacy and its role in society, in individual agency, and in technical and legal social perspectives (Martin, 2012; Toubiana et al, 2011; Turow, 2012). Related developments such as cloud storage have further complicated these issues (Stark and Tierney, 2013). At the same time, keeping track of the venues where discussion of these topics is taking place as well as the discussions themselves is difficult. Conversations about privacy are happening across academic fields, disciplines, and traditions, as well as within government and industry.

In some circles privacy and security issues go hand-in-hand, but in this work we do not explicitly account for situations in which security concerns intentionally infringe upon privacy. For example, implementations of Digital Rights Management where user behavior might be monitored, ostensibly to protect intellectual property. What privacy scholars do agree on, however, is that security is related to privacy, that myriad ethical issues exist, and it is important for those involved in information technology to stay abreast of the myriad dimensions of ethics in their work.

This study is a first step at understanding how those in academia and industry engage the topics of privacy and ethics. What overall themes, if any, exist in the literature and how are those themes inter-related? The definitions, scope, and discussion of privacy can often get complicated in the specific settings, contexts, or perspectives that are accounted for in that instance. We sought to empirically identify the themes and connections between privacy and ethics in the work being conducted across academia and industry. We focus specifically on the lexicon and discourse of privacy and ethics as they appear in published work.

To establish the scope of our study, we investigated the question of *where* experts from both academic and industry have been publishing on the subjects of privacy and ethics. To this end, we queried the Thomson Reuters Web of Knowledge (WoK) and IEEE Xplore databases using keyword pairings of engineering field and privacy concerns, shown in Table 1. The engineering field and privacy concern keywords were generated by practicing engineers and privacy researchers, respectively. These keyword lists were then jointly approved by two professors specializing in electrical and computer engineering and in

ethics, education, and communication. Our search results returned many social science, medicine, and law publications, each with only a few papers matching our search criteria. These results suggested that there are many small, isolated conversations in various fields. Of all the publication venues returned by our search, *IEEE Security & Privacy* was ranked second in matched papers and returned almost three times more papers than the next most prolific venue.

The high matching paper count is a strong indication that *IEEE Security & Privacy* is a prominent clearinghouse for the topics in Table 1. While this certainly does not imply that *IEEE Security & Privacy* is a comprehensive collection, we assume that due to its popularity amongst both authors from industry and academia that it is a representative sample of the interdisciplinary conversation regarding privacy and ethics. In fact, our analysis showed that the top ten most prolific institutions publishing articles were roughly evenly split between academic and non-academic institutions. *IEEE Security & Privacy* covers a wide range of topics and has the explicit aim of facilitating conversation across fields as well as the public about privacy—their stated goal is to disseminate research to a "general, technically-oriented readership." For these reasons, we chose *IEEE Security & Privacy* as our bibliometric analysis candidate.

Table 1 Initial Search Keywords

Engineering Field	**Privacy Concern**
Artificial Intelligence, AI, machine learning	Authentication, breach, notification
Cloud	Aware* (Wicker and Schrader, 2011)
Computer engineering, computer architecture	Consumer, marketing, advertising (Acquisti and Grossklags, 2005)
Data mining, data storage, data retrieval, natural language processing	Crypto*, encrypt*
GPS, global positioning	Education
Information science, information systems	Ethic, moral, rights (Allen, 2007)
Internet	Government, law, regulation
Medical, biotechnology, bioengineering, nanotechnology	Human-computer interaction (Knobel and Bowker, 2011)
Mobile, cellular	Protection, risk
Network theory, communications	Society (Stajano and Wilson, 2009)
RFID, radio frequency identification	Surveillance, dataveillance, tracking (Allen, 2007)
Social network	
Software	

IEEE Security & Privacy published its first issue in January 2003 with the goal of reaching a broad array of academics, engineers, and policy makers. Since then, *IEEE Security & Privacy* has emerged as an important outlet for the dissemination of research and discussion. We extracted several key words in the magazine's statement of scope as of February 06, 2014:

> "...diverse aspects of **security** and **dependability** of computer-based systems, including **legal** and **ethical** issues, **privacy** concerns, tools to help secure information, methods for development and assessment of **trustworthy systems**, analysis of **vulnerabilities and attacks**, **trends and new developments**, pedagogical and curricular issues in **educating** the next generation of security professionals, secure **operating systems** and **applications**, security issues in **wireless** networks, **design and test** strategies for secure and survivable systems, and **cryptology**, and other topics of interest to a general, technically-oriented readership." (emphasis ours)

Over the past 10 years the magazine has published on a variety of topics from a plurality of fields loosely enumerated by the words in bold above. We present an analysis of the contents of *IEEE Security & Privacy* from its first publication in 2003 to February 06, 2014, focusing on two key sets of questions:

1. What themes appear in the published academic work or "academic conversation" regarding security and privacy in *IEEE Security & Privacy* and how do these theme relate to the statement of scope? Furthermore, what are the inter-theme relationships?
2. Where is the work being conducted, and by whom?

In order to address the first question, we conducted a content or "thematic" analysis of articles published in *IEEE Security & Privacy*, deriving content "themes." These themes are not mutually exclusive, so we augmented our content analysis with JK-biplots (Torres-Salinas et al, 2013), a variant of principal component analysis (Jackson and Trochim, 2002). Finally, we determine the most prolific authors and institutions publishing in *IEEE Security & Privacy*, as well as identifying the highest cited articles. This allows us to discover the origin of work published in *IEEE Security & Privacy* as well as its relative impact.

2 Methods

Our thematic analysis of the contents of *IEEE Security & Privacy* is built on the text of publication abstracts and publication titles, similar to the methodology of (Azevedo et al, 2010). In short, we used the diction of article titles and abstracts to discover commonly covered subjects or themes in *IEEE Security & Privacy*, then classified all articles into one or more more of these themes. We built the themes through manual examination of a small set of articles, then ran the remaining articles through an automatic classification engine. This engine matched articles against themes by testing for the presence or absence of keywords associated with a particular theme in an article's abstract. An outline of our process is shown in Figure 1.

1. Manual operations on subset of dataset:
   (a) Randomly select 300 articles from the set of all articles in *IEEE Security & Privacy*.
   (b) Extract descriptive keywords from each article's title/abstract.
   (c) Generate sub-themes by grouping keywords by subject.
   (d) Generate themes by grouping sub-themes.
2. Automatic operations on full dataset:
   (a) Keyword match against full dataset.
   (b) Based on keyword match, classify each article into sub-themes/themes.
   (c) Perform JK-biplot analysis on classification results.
   (d) Obtain author, institution, and citation count metadata.

Fig. 1 Thematic Analysis Process

2.1 Manual Operations

In order to extract themes from our dataset, we randomly selected 300 articles from our dataset to manually examine. This "sampling" had the additional benefit of mimicking the Ockham's hill method to avoid over fitting to a data set (Gauch, 1993). We built a set of themes based on article content by having two individuals read the titles and abstracts of each of the randomly selected articles (Weber, 1990). During the reading, they extracted keywords that represented/described the article contents, similar to (Azevedo et al, 2010). Singular and plural forms of words were lumped as a single keyword instance, as were different hyphenation cases (e.g. "cyberwarfare" versus "cyber-warfare" versus "cyber warfare"). Different part of speech and conjugation cases such as "cryptography" versus "cryptographic" were also treated as a single keyword and resolved using wildcards, e.g. "crypto*". This process generated approximately 300 unique keywords, each of which was present in at least two different articles.

Subsequently, we categorized similar or related keywords into 35 sub-themes. For example, the keywords "firewall" and "antivirus" imply similar subject matter and are therefore are grouped along with other keywords in a "Defense" sub-theme. Further investigation revealed relationships between sub-themes, so we aggregated the 35 sub-themes by subject into a set of 8 themes. Table 2 shows the sub-theme to theme mapping, the total number of articles associated with each theme and sub-theme, as well as abbreviations for each theme. For brevity, we use these abbreviations for the rest of this article. Please note that articles can match multiple keywords, sub-themes, and themes. Table 2 also contains a mapping between our themes and the keywords extracted from the *IEEE Security & Privacy* statement of scope, indicated in bold in Section 1. Mapping the statement of scope to our themes suggests that the body of articles published in *IEEE Security & Privacy* provides full coverage of the subjects in the statement of scope. For clarification, our use of "Social Engineering" as a sub-theme, as seen in Table 2 under "Threat," follows Hadnagy's definition, which is the "...act of manipulating a person to take an action that may or may not be in the target's best interest" (Hadnagy and Wilson, 2010).

Table 2 Themes and Sub-Themes

Theme	Articles	Sub Themes (n)	Statement of Scope
Applied Context (AC)	573	Military/Defense (36), Health systems (30), Cloud (35), Infrastructure (250), Hardware (67), Software (316)	wireless, operating systems, applications
Business, Finance, and Economics (BFE)	552	Consumer (120), Economics (259), Business/Management (350), Banking/Finance (59)	trends, new developments
Cryptography (Crypto)	215	Methodology (91), Verification (141), Key systems (19)	trustworthy systems, cryptology
Data & Information Management (D&IM)	348	Big Data (105), Data protection (226), Personal identity (102)	secure information
Design (Design)	591	Interface (194), Security trade-offs (128), Real-world application (24), System design (277), Programming (82), Systems engineering (92)	dependability, tools, design, test
Privacy Awareness & Education (PA&E)	447	Education (224), Professional Environment (130), Public Awareness, Media and Communications (168)	privacy, educating
Socio-Political (SocPol)	616	Fraud and Identity Theft (29), Intellectual Property (24), Law (197), Institutional/Governmental (447), Ethics (97)	legal, ethical
Threat (Threat)	566	General threat (475), Passive Threat (11), Active threat (93), Defense (213), Social Engineering (48)	security, vulnerabilities, attacks

Articles can be matched against multiple themes and sub-themes.

During the keyword to theme/sub-theme mapping process, our list of keywords was pruned to avoid aliasing between different meanings. For example, we removed keywords such as "right" or "rights," as such a word was used in both a law or ethical sense as well as in the context of access control rights. While this does reduce the number of keywords, the overall number of keywords removed in this fashion was less than 10. At the end of pruning, we were left with 301 unique keywords grouped into 35 sub-themes. Each of the 8 themes reflects conceptualizations of security and privacy that incorporate issues addressed and characterized by *authors* published in *IEEE Security & Privacy*. The scope of thematic content ranges from technical discussions of cryptography to ethical, legal, and cultural discussions on the nature of privacy.

The remaining articles were then processed through an automatic classification engine, as described below. This engine matched articles against themes by testing for the presence or absence of keywords associated with a particular theme in the abstract or title of an article.

2.2 Automatic Operations

In the manual operations phase, we built a theme hierarchy, starting by grouping keywords into sub-themes and then grouping those sub-themes into themes. This generated a keyword-to-theme mapping, which we leveraged to generate a article-to-theme mapping for *all* articles in our dataset. In order to obtain an article-to-themes mapping, we mapped our articles to keywords, which we mapped to themes using our categorizations from Section 2.1.

We obtained an articles-to-keywords mapping by searching the title and abstract of each article for keyword occurrences using a case-insensitive regular expression engine—the automatic classification engine referred to earlier. For example, if an article abstract contains the word "virus," that article would be associated with the "Active Threat" sub-theme and the "Threat" theme. Each article can be associated with multiple sub-themes and themes, e.g. if the abstract also contained the word "law," it would also be classified under the "Law" sub-theme and the "Socio-Political" theme.

The end result of this keyword search process was an associativity matrix of size $n \times 8$, where n is the number of articles published in *IEEE Security & Privacy* up to February 06, 2014 (1690) and 8 is the number of themes. A row in this matrix represents a single article, with each row element indicating if that article is associated with a theme. A 1 indicates an article is associated with that theme and a 0 indicates no keywords associated with that theme are present in the title or abstract of the article. A total of 385 articles did not match any themes. While this represents slightly over 20% of our dataset and may seem abnormally high, 535 articles had 20 or less words in their abstract, including 84 with empty abstracts. These short-abstract articles are typically recurring magazine columns with the same one-sentence abstract for each appearance of the column in *IEEE Security & Privacy*. We did not remove these regular periodical columns or other non-article data such as book reviews, editorials, or the Table of Contents as we believe this content is relevant to the academic discussion. The fact that the number of short-abstract articles exceeds the number of articles that were not matched to a theme supports our assumption.

Table 3 presents two different measures of pairwise theme relationship. The first measure, Matching Articles, is a count of how many articles matched both themes in every possible theme pairing. The second measure is the Pearson Chi-square (χ^2 Independence Measure) (Pearson, 1900), which reports a p value. A p value less than 5% suggests we can reject the null hypothesis that two themes are independent. We use the count data represented as a contingency table to obtain the Pearson Chi-square p value, which we report as a percentage in Table 3. The results of the Chi-square test would suggest that we can reject the null hypothesis that themes are independent for all but the Crypto-PA&E pairing. In other words, we cannot differentiate between themes using just statistical tests of independence.

In order to extract multi-theme, i.e. more than just pairwise, relationship information from the article-to-theme associativity matrix, we turn to a tech-

Table 3 Theme Statistics

	AC	**BFE**	**Crypto**	**D&IM**	**Design**	**PA&E**	**SocPol**	**Threat**
AC	573	280	115	169	326	204	277	340
BFE	0.000	552	97	207	313	198	286	283
Crypto	0.000	0.003	215	95	146	61	93	100
D&IM	0.000	0.000	0.000	348	202	121	191	181
Design	0.000	0.000	0.000	0.000	591	199	276	309
PA&E	0.000	0.000	**49.393**	0.008	0.000	447	219	195
SocPol	0.000	0.000	2.645	0.000	0.000	0.000	616	273
Threat	0.000	0.000	0.001	0.000	0.000	0.000	0.000	566
	AC	**BFE**	**Crypto**	**D&IM**	**Design**	**PA&E**	**SocPol**	**Threat**

Legend: χ^2 Independence Measure (%) Matching Articles (n)

nique like principal component analysis (PCA). Specifically, we use the JK-biplot tool of (Torres-Salinas et al, 2013). In JK-biplot analysis, each article is represented by row in the associativity matrix where each element represents one of the eight themes in Table 2. As described earlier, each row element is 1 if there is a theme match and 0 otherwise. In essence, the dataset is represented by a point cloud of articles in an 8-dimensional space.

In order to collapse this space into a number of dimensions we can visualize, JK-biplot analysis uses PCA to collapse eight to two dimensions or "components." In the simplest form, the 8-dimensional point cloud is projected onto a two dimensional space where the coordinate axes were chosen to maximize the variance covered by each component, i.e. each component lies across the largest dimension of the projected point cloud.

JK-biplot analysis differs from that of PCA in that it also calculates the projection of a unit vector along each of the axes from the eight dimensional space onto the two dimensional space. We present the results of our JK-biplot analysis in Section 3.1. Interpretation of the JK-biplot is as follows:

- *Points* represent articles. The coordinate of each point is in 8-dimensional space collapsed to two dimensions. Points in close proximity to one another indicate those articles are closely related thematically.
- *Vectors* represent the original axes/themes. The angle between vectors represents the degree of inter-relationship between themes. A smaller angle indicates a higher degree of inter-theme relationship.

In addition to examining themes and their relationships, our second research question required an examination of the bibliometric metadata for the most productive authors and institutions as well as the most highly-cited articles in *IEEE Security & Privacy*. Author and institution names were extracted from the metadata available on IEEE Xplore. We used the ISI Web of Knowledge database (using their XML SOAP API) to obtain the citation count, as it is less prone to citation over-counting than services such as Google Scholar (Beel and Gipp, 2010). To determine author productivity, we made use of the Adjusted Productivity Score (APS), which credits each of the authors of an article with n authors with $1/n$ authorship (Lindsey, 1980). These bibliometric metadata results are shown in Section 3.2.

3 Results and Discussion

3.1 JK-biplot Analysis

Figure 2 shows the raw two dimensional JK-biplot data, i.e. the original $n \times 8$ associativity matrix compressed by PCA into 2 dimensions. Each dot represents an article, and each vector represents one of our 8 themes. For legibility, we scaled each vector by a factor of 4. Also, because there are only 2^8 possible coordinates, each point in the JK-biplot may represent multiple articles, i.e. many points overlap due to the binary nature of our associativity matrix. As described in Section 2.2, 385 articles did not match any themes. This creates an artifact in the PCA resulting in negative component axes, as seen in the figures below.

As is typical in PCA-transformed data, each axis or "component" is attributed a certain meaning. We colorized the dots representing each article in Figure 2 to show how many themes each article matched against. A dark dot represents an article that matched few or no themes, i.e. its title/abstract did not contain keywords associated with any themes. A lighter dot represents an article that matched many or all of the 8 themes. A clear mapping mapping from dot color to position is evidenced along Component 1, "Theme Coverage." An article that has good Theme Coverage at least touches on each theme in the contents of its title and abstract, and one with poor Theme Coverage has few keywords or keywords associated with few or even no themes.

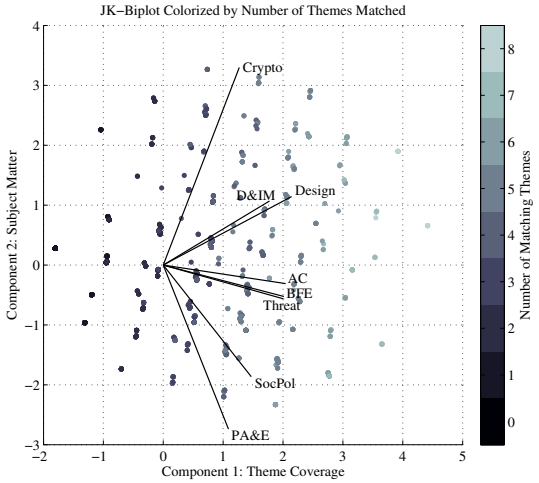

Fig. 2 JK-biplot Colorized by Number of Matching Themes

We labeled Component 2 in Figure 2 as "Subject Matter," due to the relative angles of each theme vector. Component 2 roughly maps to technical content, with Crypto, Design, and D&IM at one end and BFE, SocPol, and

PA&E at the other. Figure 3 illustrates our earlier discussion of unit vector projections. The data for each JK-biplot in Figures 2 and 3 are the same—the difference is in the dot colorization. In Figure 3, blue dots represent articles associated with a particular theme and gray dots represent articles not associated with that theme. Visual inspection suggests that the Crypto and PA&E theme vectors are directed towards the parts of the point cloud most dense with articles related to those themes. Figure 3 adds some insight into our high p value between Crypto and PA&E from Table 3. We have omitted colorized JK-biplots for the other 6 themes for brevity, but the general trend is that the theme vectors fall along the section of highest theme matching density in the point cloud. The points associated with theme vectors with angles close to 0 generally covered most of quadrants I and IV. We chose to show Crypto and PA&E as they are the most informative and illustrative of the JK-biplot technique.

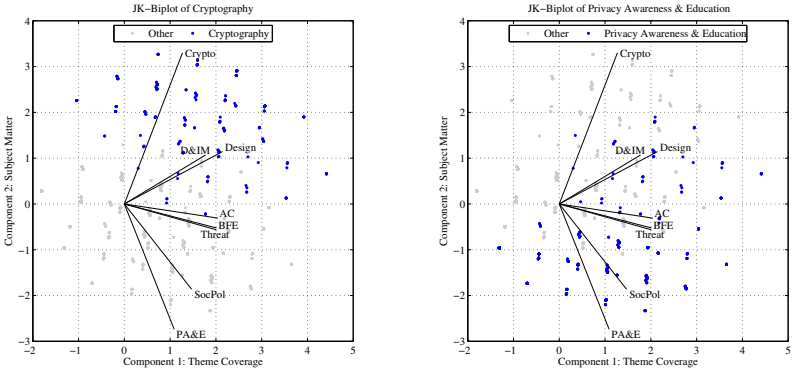

Fig. 3 Cryptography and Privacy Awareness & Education Themes

Combining the Theme Coverage and Subject Matter components together suggests several characteristics of the dataset:

- Articles focused on a specific subject theme score low on Theme Coverage (Component 1) and high (by absolute value) on Subject Matter (Component 2). Good examples of this are articles on cryptography, which tend to be highly focused and therefore typically only match the Crypto theme. This is not a byproduct of the "technical" content of the articles, but rather an indication of focus on one subject. Articles matching the PA&E theme also scored low on Component 1 and high (by absolute value) on Component 2, suggesting that articles covering subjects related to Crypto and PA&E *only* covered Crypto and PA&E, respectively.
- Articles that covered a wider range of subject matter tended to match multiple themes and have a high value x coordinate, i.e. they have a high value in Component 1, Theme Coverage.

- Themes with small angles relative to the x-axis represent themes that are often discussed in articles along with other themes. These themes are generally broader in scope and are used as motivation or as context for the work presented in articles. Good examples of this are the BFE and AC themes—work is often targeted for a specific context (AC) and/or motivated from a financial or business perspective (BFE).
- Themes with small angles relative to the x-axis do not necessarily have a high matched article count for that theme. As described above, this indicates that articles matched to a theme are also likely to be matched against other themes. For example, D&IM is relatively close to the x-axis and is the second smallest in matched article count.

This thematic analysis also sheds light on the multi-dimensional problem of the privacy conversation, where scholars necessarily have to abstract some of the complexity and focus in on a particular area. For example, Gürses and Diaz (Gurses and Diaz, 2013) have examined how there are a variety of different approaches and perspectives that scholars approach the privacy problem from, ranging from the surveillance narrative, social privacy perspective, and the practices of the industries that control private data. Useful future work includes an investigation of how our thematic mapping relates to frameworks such as that of Gürses and Diaz. As it stands, our analysis of the articles within *IEEE Security & Privacy* shows the different thematic mappings of the privacy discussion, the proportion of themes, the degree of inter-theme overlap, and indicates areas which might be understudied.

While one might expect that Crypto to be narrowly focused on its own subject matter with little theme overlap, we did not expect the (PA&E) theme would be similarly lacking in theme overlap. This is surprising for several reasons, and has important implications for how journal articles discuss and conceptualize privacy awareness. The lack of overlap indicates that PA&E is not put in the context of other issues. Based on our thematic analysis, it appears that within PA&E articles privacy is considered a general umbrella term as opposed to being linked with a more specific message about privacy situated in the context of other themes. The focus of articles matching PA&E seems to be on how to affect people's attitude/behavior/efficacy in specific contexts. While these are critical issues, thinking about privacy awareness only in these terms is limited, considering that developments in AC, BFE, Design, and D&IM have significant implications for people's awareness of threats and their ability to protect their privacy. As a concrete statement, with the increasing importance of the ethical concerns surrounding privacy, we believe that the PA&E vector should lie more towards the x-axis of our JK-biplots. This would suggest that the theme of Privacy Awareness & Education was well represented across many articles either explicitly or as a concept framing the article. This could be an area to address structurally within the larger privacy discussion.

3.2 Bibliometric Metadata

In addition to our thematic analysis, we also obtained data regarding authors and institutions productivity in this magazine and articles citation count. Table 4 shows the top ten most productive in *IEEE Security & Privacy* authors by APS, as discussed in Section 2. Of the 1413 authors in the dataset, 34 met or exceeded an APS of 5.0. Note that content with an empty author metadata field were omitted from this analysis. The arithmetic mean of APS was 0.91, so the top ten authors are at least tenfold more productive than the average. Of the top ten, 4 are university-affiliated and 6 are industry or IEEE affiliated, suggesting that both academia and industry are well represented among the most prolific *IEEE Security & Privacy* authorship.

Also in Table 4 are the top ten institutions by number of articles published. The affiliation of the first author of each article determined which institution received credit. As with the authors, we were unable to determine the appropriate affiliation for some articles and they have been omitted from this portion of the analysis. Of the 376 known affiliations, only 15 institutions met or exceeded an article count of 10, but 43 institutions have published at least 5 articles. The average number of articles published across all 376 institutions was 2.48. As with the top ten authors, the top ten institutions span both academia and industry, with 6 universities and 4 industrial institutions in the top ten.

Table 4 10 Most Productive Authors and Institutions in *IEEE Security & Privacy*

APS	Author	Affiliation
54.78	McGraw, G.	Cigital
35.00	Geer, D.E.	@stake, In-Q-Tel, Verdasys
32.00	Lesk, M.	Rutgers University, Internet Archive
27.00	Donner, M.	Morgan Stanley, Associate Editor in Chief, Google
25.50	Schneier, B.	BT, Counterpane Internet Security
18.04	Bellovin, S.	Columbia University
17.00	Garber, Lee	IEEE Computer Society
16.50	Ortega, B.	IEEE Computer Society
15.50	Schneider, F.	Cornell University, Associate Editor in Chief
14.39	Pfleeger, S.L.	Dartmouth College, RAND

Papers	First Author Affiliation
53	Cigital
38	Dartmouth College
27	In-Q-Tel
25	Columbia University
25	Rutgers University
22	Carnegie Mellon University
17	Microsoft
13	BT
13	UC Davis
13	Cornell University

Separating institutions into Academic and Non-Academic categories shows that, at least based on the first author affiliation, both academia and industry are well-represented in *IEEE Security & Privacy*. On average, academic institutions submitted 2.29 articles compared to 2.75 articles for non-academic institutions. 20 academic institutions have at least five articles with 23 non-academic institutions also publishing at least five articles in *IEEE Security &*

Privacy. Table 5 shows the article count for the top ten institutions in both categories.

Table 5 10 Most Productive Institutions by Type

Academic		Non-Academic	
Papers	**First Author Affiliation**	**Papers**	**First Author Affiliation**
38	Dartmouth College	53	Cigital
25	Columbia University	27	In-Q-Tel
25	Rutgers University	17	Microsoft
22	Carnegie Mellon University	13	BT
13	UC Davis	13	IBM
13	Cornell University	12	Google
11	Florida Inst. of Tech.	11	Core Security Technologies
8	Indiana University	10	RAND
8	Pennsylvania State University	8	Counterpane Internet Security
8	US Naval Postgraduate School	8	SRI International
8	University of Maryland	8	Verdasys

Of the 1690 articles in *IEEE Security & Privacy*, only 427 had at least one citation, according to the ISI Web of Knowledge. Of the 427 articles, the average citation count was 7.15. 160 had more than 5 citations and 80 had more than 10. The top ten cited articles in *IEEE Security & Privacy*, shown in Table 6, exceed the average by over tenfold in most cases.

Table 6 10 Most Cited Articles in *IEEE Security & Privacy*

Cites	**Title**	**Authors**
148	The security and privacy of smart vehicles	Hubaux, J.P., *et al.*
112	RFID privacy: an overview of problems and proposed solutions	Garfinkel, S.L., *et al.*
98	A survey of secure wireless ad hoc routing	Yih-Chun, Hu. and Perrig, A.
90	Security and Privacy Challenges in the Smart Grid	McDaniel, P. and McLaughlin, S.
82	Secret-ballot receipts: True voter-verifiable elections	Chaum, D.
81	The spread of the Witty worm	Shannon, C. and Moore, D.
69	Privacy and rationality in individual decision making	Acquisti, A. and Grossklags, J.
67	Password memorability and security: empirical results	Yan, J., *et al.*
65	Toward Automated Dynamic Malware Analysis Using CWSandbox	Willems, C., *et al.*
52	Smart-grid security issues	Khurana, H., *et al.*
50	Overview of IEEE 802.16 security	Johnston, D. and Walker, J.

4 Conclusion

Our bibliometric data analysis shows that the most prolific authors and institutions of *IEEE Security & Privacy* come from both academia and industry, suggesting that *IEEE Security & Privacy* is the venue for both academics and those practicing in industry to exchange ideas. From our JK-biplot analysis, it seems that we have two relatively "disconnected" areas of investigation

within *IEEE Security & Privacy*. The two themes are PA&E and Crypto, as they are the farthest in angle not only from the x-axis but also from all other themes. The relative thematic distance of Crypto is understandable, as it is a specific, technical subject. However, we believe future articles on the theme of Privacy Awareness & Education (PA&E) could be more tightly integrated with other themes. The position of BFE in our JK-biplot suggests that many articles are motivated by or at least touch on the subject of business and economic interests. This is in stark contrast to the PA&E vector, which is almost perpendicular to our Theme Coverage component (Component 1).

This apparent emphasis on business and economic interests as opposed to privacy awareness and education is not a negative trait, but rather represents a research space in which to expand. On the business side, corporations have established a valuation system for privacy, targeted web ads being one example, but the question of the value of privacy to the individual is only now being explored (Acquisti, 2014; Acquisti et al, 2013; Schrader et al, 2013). Market forces are quite important in driving the progress of technology, but as a community we should give some if not equal attention to the ethical considerations of technology in this age of data. While there has been work on connecting business interests with that of the consumer, especially with smart grid power monitoring systems (Ghosh et al, 2012), we believe a more generalized set of standards, laws, and engineering practices should be implemented to address privacy ethics and awareness.

We are moving forward into an age where big data and data mining enable privacy violation. Privacy and ethics education are paramount for both the general public as well as engineers, social scientists, information technologists and educators. The importance of adequate consideration of ethical implications in technical work and the importance of acknowledging and closing the present gaps in addressing ethical issues in the study of privacy and security will only increase with time. *IEEE Security & Privacy* is only one venue where discussions of the ethical and technical aspects of privacy are occurring, but it is a dominant one with a wide area of focus. We hope this work encourages everyone, particularly researchers in the field of privacy and security, to take a proactive role in the ethical considerations of their work as well as their publications.

Acknowledgments

The authors would like to acknowledge Jubo Yan, William Schulze, and Stephen B. Wicker of Cornell University for their invaluable contributions. This research was funded by grant 1016203 from the National Science Foundation and there are no conflicts of interest by any author or consultant to this project.

References

Acquisti A (2014) From the Economics of Privacy to the Economics of Big Data. Privacy, Big Data, and the Public Good: Frameworks for Engagement

Acquisti A, Grossklags J (2005) Privacy and rationality in individual decision making. IEEE Security & Privacy 3(1)

Acquisti A, John LK, Loewenstein G (2013) What Is Privacy Worth? http://dxdoiorg/101086/671754 42(2):249–274

Allen A (2007) The virtuous spy: privacy as an ethical limit. Scholarship at Penn Law

Azevedo PG, Mesquita FO, Young RJ (2010) Fishing for gaps in science: a bibliographic analysis of Brazilian freshwater ichthyology from 1986 to 2005. Journal of Fish Biology 76(9):2177–2193

Beel J, Gipp B (2010) Academic Search Engine Spam and Google Scholar's Resilience Against it. Journal of electronic publishing

Cohen JE (2013) What Privacy is For. Harvard Law Review 126

Diebold FX (2003) Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress . Econometric Society Monographs, Cambridge University Press, Cambridge

Gauch H Jr (1993) Prediction, Parsimony and Noise. American Scientist 81(5):468–478

Ghosh DP, Schrader DE, Schulze WD, Wicker SB (2012) Economic analysis of privacy-aware Advanced Metering Infrastructure adoption. Innovative Smart Grid Technologies (ISGT), 2012 IEEE PES pp 1–4

Gurses S, Diaz C (2013) Two tales of privacy in online social networks. Security & Privacy, IEEE 11(3):29–37

Gürses S, Troncoso C, Diaz C (2011) Engineering Privacy by Design. ... Privacy & Data Protection

Hadnagy C, Wilson P (2010) Social Engineering: The Art of Human Hacking, 1st edn. Wiley Publishing, Indianapolis

Hugl U (2010) Approaching the value of Privacy: Review of theoretical privacy concepts and aspects of privacy management. AMCIS

Hull G, Lipford HR, Latulipe C (2011) Contextual gaps: privacy issues on Facebook. Ethics Inf Technol

Jackson K, Trochim W (2002) Concept mapping as an alternative approach for the analysis of open-ended survey responses. Organizational Research Methods 5(4):307–336

Knobel C, Bowker GC (2011) Values in design. CACM

Lessig L (2006) Code. Lawrence Lessig

Lindsey D (1980) Production and Citation Measures in the Sociology of Science: The Problem of Multiple Authorship. Social Studies of Science 10(2):145–162

Martin K (2012) Information technology and privacy: conceptual muddles or privacy vacuums? Ethics Inf Technol 14(4):267–284

Pearson K (1900) X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it

can be reasonably supposed to .... The London
Pham MC, Klamma R, Jarke M (2011) Development of computer science disciplines: a social network analysis approach. Soc Netw Anal Min 1(4):321–340
Richards NM, King JH (2014) Big Data Ethics. Wake Forest Law Review
Sar RK, Al-Saggaf Y (2014) Contextual integrity's decision heuristic and the tracking by social network sites. Ethics Inf Technol
Schrader D, Ghosh DP, Schulze WD, Wicker SB (2013) Civilization and Its Privacy Discontents. In: Patterson H (ed) Privacy Law Scholars Conference, Berkeley, CA
Shilton K (2010) Participatory Sensing: Building Empowering Surveillance.: EBSCOhost. Surveillance & Society
Solove DJ (2007) The Future of Reputation: Gossip, Rumor, and Privacy on the Internet. Yale University Press
Stajano F, Wilson P (2009) Understanding scam victims: seven principles for systems security. Tech. Rep. UCAM-CL-TR-754, University of Cambridge
Stark L, Tierney M (2013) Lockbox: mobility, privacy and values in cloud storage. Ethics Inf Technol 16(1):1–13
Torres-Salinas D, Robinson-Garcia N, Jiménez-Contreras E, Herrera F, López-Cózar ED (2013) On the use of Biplot analysis for multivariate bibliometric and scientific indicators. arXiv 1302.0608v1
Toubiana V, Narayanan A, Boneh D, Nissenbaum H (2010) Adnostic: Privacy Preserving Targeted Advertising. NDSS
Toubiana V, Subramanian L, Nissenbaum H (2011) TrackMeNot: Enhancing the privacy of Web Search. arXiv 1109.4677v1
Turow J (2012) The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth. Yale University Press
Weber RP (1990) Basic Content Analysis, 2nd edn. SAGE Publications, Newbury Park, CA
Wicker SB, Schrader DE (2011) Privacy-Aware Design Principles for Information Networks. Proc IEEE 99(2):330–350