IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 9, 2020, accepted August 19, 2020, date of publication August 25, 2020, date of current version September 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019334

A Study of South Asian Websites on Privacy Compliance

YOUSRA JAVED[1], KHONDAKER MUSFAKUS SALEHIN[2], (Senior Member, IEEE), AND MOHAMED SHEHAB[3]

[1] Department of Computing, National University of Sciences and Technology, Islamabad 46000, Pakistan
[2] Department of Computer Science, California State University, Dominguez Hills, CA 90747, USA
[3] Department of Software and Information Systems, University of North Carolina, Charlotte, NC 28223, USA

Corresponding author: Yousra Javed (yousra.javed@seecs.edu.pk)

ABSTRACT Privacy laws in South Asian countries are still at a nascent stage. Therefore, South Asian websites are susceptible to user privacy violation. This paper presents an assessment of website privacy policies from 10 sectors in the three largest South Asian economies, namely, India, Pakistan, and Bangladesh. Using a manual qualitative analysis on a dataset of 284 popular websites, we assessed the policies based on accessibility, readability, and compliance with 11 privacy principles. Our findings show that overall, the privacy statement accessibility, and privacy compliance of websites from the three countries is low especially in the education, healthcare, and government sectors. Readability is quite low for websites in all 10 sectors of the three countries. Privacy compliance in each country is the highest for the principles of data processing and third-party transfer, whereas it is the lowest for protection of children's data, data retention and portability. Indian websites performed comparatively better amongst the three countries on all three metrics, followed by Pakistan, and Bangladesh. Based on our results, we provide recommendations involving all stakeholders (i.e., website owners, privacy regulators, and users) to help improve privacy protection of user data in South Asia.

INDEX TERMS Accessibility, data protection, GDPR, privacy compliance, privacy law, readability, web.

I. INTRODUCTION

Privacy is the ability of an individual to express himself selectively in a public domain [1], [2]. In this digital age, the notion of privacy primarily refers to the freedom that an individual should have for determining how his personally identifiable information (e.g., name, date of birth, email address, and IP address) are processed, i.e., collected, used, and disclosed [3]–[5].

One channel through which personal information is processed, is the publicly accessible websites of business and service entities. Ensuring that the users of these websites are aware of how their personal information is processed, how the accuracy of their data is maintained, how its data integrity and confidentiality is preserved. is vital. These aspects are important because user information is now considered a valuable commodity. For example, business entities either analyze user information themselves or share/sell them to advertisers and researchers to best tailor commercial services to the online consumer market. Therefore, it is expected that all websites include a relevant privacy statement [17] to ensure lawful, fair, and transparent processing of user data [6].

A website's privacy policy needs to comply with its regional privacy and data protection laws. These laws originated from the Fair Information Practice Principles (FIPPs) outlined by the United States Federal Trade Commission [8]. Lately, new and more focused privacy legislations have been enacted regionally that fit to the local privacy requirements in the region. The General Data Protection Regulation (GDPR) is one such legislation created by the European Union (EU) [6]. GDPR is globally considered as the most comprehensive and the strongest set of privacy and data protection laws [7]. Failure to comply with its guidelines by a website that is EU-based or offers goods/services to a user in the EU can result in a sizable fine. For instance, Google was fined €50 million in 2019 because it failed to provide clarity and transparency on how it handled personal data of its French users for targeted advertising [9].

The concept of digital privacy is a modern construct, primarily associated with the western world; it has remained

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang.

virtually unknown in some regions until recently. Amongst them, South Asia is an acute example. The three largest South Asian economies, namely, India, Pakistan, and Bangladesh [10], introduced their first privacy and data protection laws in the last two years. Since the implementation of user privacy is still at a nascent state in these three countries, their websites are susceptible to user privacy violation. Two recent examples corroborate this statement: i) a healthcare website in India failed to protect data of its 6.8 million users from leakage [11] and ii) a large e-commerce business and a leading ride-sharing company in Bangladesh were found to collect user data without their consent [12].

It is, therefore, important to study whether South Asian websites have easily accessible and understandable privacy statements for their users and whether these statements comply with the established privacy and data protection laws. Although similar studies on US, European, and some Middle Eastern websites are available [13], [15], [16], no such study concerning South Asian websites is available to the best of our knowledge. We take a first step to fill this gap with the goal of having the following three impacts: i) enable websites to improve their privacy-policy statements, ii) create awareness about privacy among their users, and iii) assist privacy regulators in improving existing privacy laws concerning best practices.

This paper performs a qualitative assessment of website privacy policies from 10 different sectors in India, Pakistan, and Bangladesh. We use GDPR as the standard for this assessment since the data protection laws of the three countries under investigation are subsets of this regulation's principles. Using a dataset of 284 popular websites, we investigate the following research questions (RQs) in the context of India, Pakistan, and Bangladesh:

RQ1: How accessible are their privacy policy statements?
RQ2: How readable are their privacy policy statements?
RQ3: How compliant are their privacy policy statements with the GDPR?

Our results show that overall, the privacy statement accessibility, and privacy compliance of Indian, Pakistani, and Bangladeshi websites is low especially in the education, healthcare, and government sectors. Readability is quite low for websites in all 10 sectors of the three countries. Privacy compliance in each country is the highest for the principles of data processing and third-party transfer, whereas it is the lowest for protection of children's data, data retention and portability. Indian websites performed comparatively better amongst the three countries on all three metrics, followed by Pakistan, and Bangladesh.

The remainder of the paper is organized as follows. Section II discusses privacy laws in South Asia. Section III summarizes existing literature related to our work. Section IV outlines our methodology for assessing accessibility, readability, and content compliance by the South Asian websites. Section V presents our results. Section VI discusses the impacts and limitations of our work. Section VII concludes the paper.

II. PRIVACY LAWS IN SOUTH ASIA

Privacy policy is a statement that discloses how an organization collects, uses, and manages its users' data [17]. It is considered as a legal document that adheres to the core principles of user privacy protection, known as FIPPS. An organization's website is required to contain a clearly stated privacy policy that is easily visible on its homepage, usually via a link labeled as "Privacy", "Privacy Policy", "Privacy Statement", etc. in the footer section.

Although privacy policy of a website generally adheres to FIPPS, this set of principles is considered arcane as they have failed to effectively enforce privacy on the Internet [8]. Therefore, modern privacy and data protection laws at regional levels have been build upon FIPPS. The content and format of a privacy policy must adhere to the laws applicable to the region or country where the corresponding website is based.

Digital privacy is a concept that is primarily associated with the western world. For example, privacy is considered as a fundamental right for a sustained democracy in Europe nowadays [18]. In contrast, South Asia is relatively new to implementing privacy laws in its constituent countries. We discuss the local privacy laws of the three largest South Asian economies along with their comparison with GDPR below.

A. INDIAN PERSONAL DATA PROTECTION BILL (IPDPB)

India is leading its effort in implementing data privacy of its citizens through a legislation bill, known as the Personal Data Protection Bill [19]. This bill is yet to be enacted as a privacy law, but the latest revision has been presented to the Indian Parliament in December 2019 [20]. A summary of its privacy principles is presented below.

i. **Data collection:** Outlines what data is collected, by whom, and for what purposes [IPDPB: Cl. 4–6].
ii. **Data retention:** Outlines the length of data storage and the corresponding review policy [IPDPB: Cl. 9].
iii. **Collection notification:** Outlines the required clarity in policy notification across different languages [IPDPB: Cl. 7].
iv. **Data sharing:** Outlines how data is shared and how data quality is maintained with third parties, including the government [IPDPB: Cl. 17].
v. **Sensitive data processing:** Outlines how sensitive data of children aged below 18 is collected and shared with parental consent [IPDPB: Cl. 16].
vi. **User control:** Outlines how users can withdraw consent to data collection, update and erase collected data, etc. [IPDPB: Cl. 18, 20].
vii. **User access:** Outlines how users would request their data and receive a copy in an accessible format [IPDPB: Cl. 19].
viii. **Security standards:** Outlines the necessity of security safeguards, their review policy, and how security breaches are notified to users [IPDPB: Cl. 24].
ix. **Quality control:** Outlines how complete and accurate the data is using a data-trust score [IPDPB: Cl. 8].

x. **Grievance redressal:** Outlines how user complaints are mitigated [IPDPB: Cl. 32].
xi. **Accountability:** Outlines the vested responsibility on the data collector for complying with the personal data protection in India [IPDPB: Cl. 10].

B. PAKISTANI PERSONAL DATA PROTECTION BILL (PPDPB)

Pakistan introduced its privacy legislation, as the Personal Data Protection Bill in 2018. An updated draft of the bill was proposed in April, 2020 [21]. This bill is expected to go through public consultations for further refinements [22]. A summary of the bill's current draft is presented below.

i. **Personal data processing:** Outlines what data is collected, and how the collected data is used [PPDPB: Cl. 5, 28].
ii. **Rights of data subject:** Outlines how users can maintain control over the sharing, collection, processing, and deletion, of their data [PPDPB: Cl. 16, 19, 23, 25-27].
iii. **Data Retention:** Outlines how the data is stored and that the processed data shall not be kept longer than is necessary for the fulfillment of the purpose [PPDPB: Cl. 9, 11].
iv. **Security requirements:** Outlines the security measures for protecting user data from any loss, misuse, modification, unauthorized or accidental access or disclosure, alteration or destruction [PPDPB: Cl. 8].
v. **Compliance with data subject's requests:** Outlines how the data controller is complying with data subject's requests regarding data access, correction, deletion, and halting processing and disclosure [PPDPB: Cl. 17-18, 20-22].
vi. **Personal data breach notification:** Outlines the requirements of notifying the users in case of a personal data breach [PPDPB: Cl. 13].
vii. **Notice to data subject:** Outlines that a written notice is provided (in English and National language) to inform the data subject regarding data collection and processing, along with the rights of the subjects [PPDPB: Cl. 6].
viii. **Data integrity and access:** Outlines that the collected personal data is accurate and that the user shall be given access to a copy of his/her personal data in an intelligible form [PPDPB: Cl. 10, 16].
ix. **Complaint:** Outlines that an aggrieved person may file a complaint in case of violation of their personal data protection rights [PPDPB: Cl. 45].
x. **Disclosure of personal data:** Outlines what data is shared, why it is shared, and whom (including govt/private sectors) it is shared with [PPDPB: Cl. 7, 14-15, 24].

C. DIGITAL SECURITY ACT (DSA)

Bangladesh implemented data privacy through a recent legislation known as the Digital Security Act [23]. Even though the enforcement of this act is already in place since October, 2018, its scope is limited to the collection process of user data. This act only outlines what constitutes personal data and how it should be collected with user consent on the Internet [DSA: Cl. 26].

D. A COMPARISON WITH THE EUROPEAN UNION

The GDPR is a European legislation on privacy and data protection. It is applicable in the the European Economic Area (EEA [24]). GDPR has been implemented since May 25, 2018; and is considered as the world's most comprehensive and strongest set of data protection laws. It outlines how people can access information about them and places limits on what business and service entities can do with personal data. We have summarized GDPR as the following 10 principles.

i. **Data processing:** Outlines that relevant personal data is processed (i.e., collected, retrieved, and used) strictly for intended purposes to which a data subject (i.e., an identified or identifiable natural person) gives consent [GDPR: Art. 5 (1)(a)-(c), Art. 6].
ii. **Protection of children's data:** Outlines that the personal data of a child below age 16 is authorized for lawful processing through parental consent [GDPR: Art. 8, Recital 38].
iii. **Third-party transfer:** Outlines that the flow of personal data to a third country or organization takes place only if the conditions for data protection and safeguards are met [GDPR: Art. 44-50, Recital 101].
iv. **Transparency:** Outlines that the data controller takes necessary measures to provide all information (e.g., policy changes and updates) referring to the personal data of data subjects, using clear and plain language [GDPR: Art. 12, Recital 39].
v. **Data retention:** Outlines that the personal data is not stored longer than what is necessary for the specified purpose(s) [GDPR: Art. 5 (1)(e)].
vi. **Data Accuracy and control:** Outlines that all reasonable steps (e.g., data subject's right to access, rectify, and delete) are taken to ensure that the user data is accurate and up-to-date [GDPR: Art. 5 (1)(d), Art. 15–17, Recital 63-66].
vii. **Right to object:** Outlines that data subject has the right to halt data processing for direct marketing purposes [GDPR: Art. 24, Recital 70].
viii. **Data portability:** Outlines that data subject has the right to receive personal data in a structured, commonly used, and machine-readable format [GDPR: Art. 20, Recital 68].
ix. **Integrity and confidentiality:** Outlines that appropriate technical or organizational measures are in place to ensure security against unlawful processing, accidental loss, and damage of personal data [GDPR: Art. 5 (1)(f)].
x. **Accountability:** Outlines that the data controller (a person or legal entity that performs data processing) is accountable for personal data processing and is

TABLE 1. A comparison between regional privacy laws in South Asia and GDPR.

Regional laws	GDPR principles
IPDPB, PPDPB, DSA	Data processing
IPDPB, PPDPB	Third-party transfer Transparency Data retention Data accuracy and user control Data portability Integrity and confidentiality Right to object Accountability
IPDPB	Protection of children's data

required to notify the supervisory authority and data subject without undue delay in case of compromised data integrity and confidentiality [GDPR: Art. 33, Art. 34, Recitals 85-88].

Table 1 presents a comparison between privacy laws in South Asia and the EU. Because GDPR is the most comprehensive privacy legislation, it is evident from this table that the proposed IPDPB (India) and PPDPB (Pakistan) are strongly following the EU guidelines for protecting personal data in India and Pakistan, respectively. However, there is a significant disagreement between DSA (Bangladesh) and GDPR concerning the implemented privacy principles. It suggests that DSA is incomplete as a privacy legislation and falls short of protecting personal data of users in Bangladesh in its current state.

III. RELATED WORK

Several studies have analyzed the privacy policies of websites across different sectors to assess their compliance with FIPPS, regional laws, and other criteria. We briefly discuss some of these studies and how we build upon this literature.

A. COMPLIANCE WITH FIPPs

Alhamod *et al.* have analyzed the privacy policies of 54 government websites of Saudi Arabia using the FIPPs guidelines [16]. Their results showed that 40% of the websites that provide privacy policy complied with two or less out of the five core privacy principles outlined in FIPPs. Similarly, Dias *et al.* analyzed 308 Portuguese government websites. They showed that only 4% of the websites that provided a privacy policy complied with FIPPs pertaining to the clarity of collection and purpose of data sharing with third parties [30]. Rains *et al.*, on the other hand, studied 97 different health websites to show that only 3% of them provided a privacy policy that met all of the FIPPs guidelines [26]. Another study by Liu *et al.* on e-commerce websites of Global 500 companies, examined compliance with FIPPs [27]. They showed that the investigated privacy policies mostly complied with the data collection principle but failed to address other aspects (e.g., user access to data) of the privacy law.

B. COMPLIANCE WITH REGIONAL LAWS

Bowers *et al.* studied the privacy policies of mobile-money services across multiple countries [15]. This study is based on the privacy principles implemented by both a regional standard (i.e., FDIC's Privacy Rule Handbook [28]), and an industry standard (i.e., GSMA's Mobile Privacy Principles [29]). They studied around 100 services and found that 50% of the websites with a privacy policy are not transparent to their users about the data collection process. Regarding government websites, Beldad *et al.* studied the privacy policies of Dutch government websites to assess their availability and compliance with the Dutch Personal Data Protection Act [30]. Their results demonstrate poor availability of privacy policy on the websites and weak compliance with the local privacy standard. On the other hand, Kuzma *et al.* analyzed 90 online pharmacy websites across nine different countries in Europe, Asia, and North America [31]. They show that the level of user data protection and privacy compliance is often very low even in those countries that have implemented strong privacy laws.

C. COMPLIANCE WITH CUSTOM CRITERIA

Desai *et al.* focused on e-commerce websites based in the US and Europe to investigate how the communication of websites' privacy policies to customers has changed over a ten year period [13]. The authors designed a policy-rating score (based on privacy content, accessibility, security, etc.) without using any established privacy legislation. Their findings based on a dataset of 525 websites show that e-commerce companies do not necessarily inform their customers of their privacy practices. More importantly, this privacy practice remained static among these companies between 2000–2010. Recently, Zaeem *et al.* studied privacy policies of 600 websites across different sectors in North America [14]. The authors first designed a survey to identify the top ten privacy concerns among users, and then manually analyzed the websites' privacy policies based on the outcomes of the survey. The analysis of these websites suggests that the majority of them shared user data with the law enforcement, and the users have limited control over their personal data following the data collection process. In another work, Robles-Estrada *et al.* analyzed the privacy statements of 120 Mexican companies in reference to the responses from a user survey [32]. Their findings show that online companies in Mexico hardly respect their customers' privacy.

Although a considerable amount of work on privacy compliance has been done as of today, no study regarding websites in South Asia is available to the best of our knowledge. South Asia is an emerging economy for a population of around 2 billion [33] and the importance of data privacy has started to gain traction in that region lately, as exemplified by the recent data breaches in this region [11], [12]. Moreover, India has declared data privacy as a constitutional right of its citizens through a Supreme Court verdict in 2017 [34]. Therefore, a study on the privacy compliance of websites based in South Asia demands a strong consideration.

IV. METHODOLOGY

This paper performs a qualitative assessment of the website privacy policies of businesses/services in the top three

TABLE 2. Number of websites per sector from the three South Asian countries.

Sector	India	Pakistan	Bangladesh
1. E-commerce[a]	10	10	10
2. Finance	10	10	10
3. Education	10	10	10
4. Healthcare	10	10	10
5. News	10	10	10
6. Government	10	10	10
7. Telecom	4	5	5
8. Buy & Sell[b]	10	10	10
9. Jobs/Freelance	10	10	10
10. Blog/Forum	10	10	10
Total (284)	**94**	**95**	**95**

[a] E-commerce refers to online shops where individual users buy products from vendors.

[b] Buy & Sell refers to online marketplaces where individual users buy and sell products/services.

economies of South Asia: India, Pakistan, and Bangladesh. We address three research questions concerning policy accessibility (RQ1), readability (RQ2), and privacy compliance (RQ3) of the selected websites. We conducted this study between March 2020 and July 2020.

Note that the first two authors contributed equally in every phase of this research whereas the third author contributed to the methodology of this work. In addition, a student verified the accessibility and readability scores of the websites.

In the following discussion, we first introduce our research dataset, then outline the individual methodologies used for answering the three RQs.

A. WEBSITE DATASET

Our dataset consists of 284 websites from 10 business/service sectors in India, Pakistan, and Bangladesh, as shown in Table 2. We chose these sectors after inspecting popular websites in the aforementioned countries according to Alexa Internet [35]. We have listed the names of the websites along with a reference for accessing the full dataset from a data repository in Appendix A.

Table 2 shows that we have selected the 10 most popular websites (or a comprehensive list of websites when the number is below 10) from each sector in the three countries. These websites were selected after reviewing information from various sources, such as Alexa Internet, Webometrics [36], Wikipedia, Quora [37], and online blogs, etc. depending on the sector. For instance, we used Webometrics website to identify the top 10 hospitals and educational institutions in each country.

B. RQ1: ACCESSIBILITY

We evaluate the accessibility of a website's privacy policy using a scoring rubric presented in Table 3. This table provides a quantifiable measure for a website on a 0–5 scale to determine whether it contains a privacy policy statement and if so, how easy is it for a user to track it on the website. For example, if the privacy statement is clearly visible on the landing page as a link (e.g., Privacy Policy, Privacy Statement, and Policy) in the footer section, the accessibility score is 5. On the other hand, the score is 0 if the website does not have a privacy statement or when the privacy statement is empty. For evaluating our dataset, we first compute the accessibility scores of all websites, then summarize them per sector from each country.

C. RQ2: READABILITY

We evaluate the readability of a privacy statement to assess how easy is it for a user to understand the privacy policy of a website. It is an important assessment criterion because privacy laws require clear statement of privacy policies in plain language [6]. For evaluating the readability of our dataset, we use Flesch-Kincaid grade-level formula, which is widely used to assess legal texts [39], medical documents [40], and privacy statements of websites [15]. This formula determines the required US grade-level education to understand a piece of text by a reader. A lower grade-level score, hence, means that the text is easier to comprehend for the reader. We calculated the Flesch-Kincaid grade-level scores using a publicly available online tool [41]. This tool also computes other readability metrics, such as Coleman Liau index, ARI (Automated Readability Index), and SMOG index. We also leveraged the basic text statistics generated by this tool such as word count of a policy statement.

For the purpose of qualitative evaluation, we calculated the average reading grade-level scores of the privacy statements in each sector to present a country-wide comparison of the readability scores. In addition, we investigated the length of each privacy statement using the word count values in this evaluation and correlated this value with the corresponding reading grade-level score. Note that word count is an important metric because the length of a privacy policy is a factor behind a user's willingness to spend time reading it [42].

GDPR requires that a privacy policy is presented to users in a clear and plain language [GDPR: Recital 39]. We used language to further analyze the readability of a privacy statement and evaluate whether it accommodates users who are familiar with languages other than English, as stated in Pakistan's data protection bill [PPDPB: Cl. 6.3]. Moreover, the three countries under investigation do not use English as their national language, and both India and Pakistan have several regional languages. Therefore, we calculated the language distributions of the website privacy policies in each country concerning English and national/regional languages.

D. RQ3: PRIVACY COMPLIANCE

We analyze the content of each privacy statement to evaluate whether the websites in our dataset comply with their regional privacy laws. In this context, we used GDPR as the reference privacy law for two reasons. First, privacy legislation in both India and Pakistan is currently at the review phase whereas the existing privacy law in Bangladesh is significantly inadequate. Secondly, the privacy laws of these three countries are subsets of GDPR principles, as shown in Table 1. In brief,

TABLE 3. Accessibility score rubric for a website's privacy policy.

Level of Accessibility	Score
Not accessible	0
Accessible through internal website search or google search	1
Policy on the Frequently Asked Questions page	2
Policy on Customer Service, Disclaimer, Terms & Conditions, or About Us pages	3
Policy on the homepage but hidden in a drop-down menu	4
Policy on the homepage, clearly visible as Privacy Policy, Privacy Statement, Policy, etc.	5

TABLE 4. Keywords used for assessing privacy compliance.

Privacy principle	Keywords
Data processing (*collection*)	collect, gather, personal, information, account, data, name, IP, mail, cookie, phone, mobile, birth, address, history, page, browser, credit, debit, payment, bank, employment, education, id, identity, survey
Data processing (*purpose*)	use, purpose, aggregate, product, service, business, ad, market, experience, promotion, improv, research
Protection of children's data	child, minor, 13, 16, 18, guard, parent
Data accuracy and control	opt, choice, choose, manage, update, delete, control, edit, withdraw, change
Data retention	retain, retention, store, duration, period, long, time, archive, expir
Integrity and confidentiality	secur, encrypt, protect, confidential, integrity, measure
Accountability	breach, attack, noti, accountab, inform, compl, law, regulation
Transparency	change, update, time, effective date, last updated, noti, modif, mail
Data portability	view, access, receive, copy, structured, readable, interoperable, PDF, format
Right to object	object, stop, grievance, dispute, complain, question, contact, mail
Third-party transfer	share, sell, sold, disclose, third, party, partner

we used GDPR for an objective analysis of privacy compliance across the three South Asian countries.

To evaluate the privacy statements for compliance, we used a keyword-based manual content analysis [15]. Table 4 lists a set of keywords that we selected to best represent each GDPR principle summarized in Section II-D. Note that we have split the data processing principle into two sub-principles: *collection* and *purpose* in the table. The notion here is to evaluate the implementation of the data processing principle in the privacy statements by separately analyzing what data items websites collect and for what specific purposes.

We manually analyzed the dataset using our keywords because automation of this process is a separate problem which is beyond the scope of this work. We first searched the content of a privacy statement using the keywords selected for each principle. If any of the keywords were found for a principle, we reviewed the texts around the matched keywords to decide whether the texts are in sufficient agreement with the definition of that privacy principle. After repeating this process for all GDPR principles, we calculated a privacy policy's compliance score as the number of principles (out of 11) that are implemented in the privacy statement.

V. RESULTS

This section presents our analysis's results. We have segmented them into 1) accessibility scores, 2) readability scores, and 3) compliance scores to answer RQ1, RQ2, and RQ3, respectively.

A. ACCESSIBILITY SCORES

Overall, 103 out of the 284 (36.2%) websites in our dataset do not have a privacy policy. Figure 1 shows an analysis concerning South Asian websites that do not provide a privacy statement for their users. The x-axis represents the website sectors and the y-axis represents the percentage of websites that have an accessibility score of 0. The bars encoded in three different patterns (i.e., diagonal, horizontal, and boxed, for India, Pakistan, and Bangladesh, respectively) summarize the analysis for each sector. In addition, the three horizontal lines (i.e., solid, dashed, and dotted) individually encode the average percentage scores for the respective countries across all sectors.

According to Figure 1, the unavailability of privacy statements on websites is consistently high across all sectors in South Asia. This observation is more evident for education (∼86%), healthcare (∼67%), government (∼57%), and blog/forum (∼50%) websites. However, we do see exceptions in the e-commerce (∼3%), news (∼10%), and buy & sell (∼13%) websites, probably, because of the popularity and volume of users in these sectors. In terms of regional analyses, around 26%, 34%, and 48% websites across all sectors in India, Pakistan, and Bangladesh, respectively, do not contain a privacy statement, as the horizontal lines show in the figure. These values suggest that the websites based in India are comparatively more transparent about sharing data privacy policies with their users.

We present a summary of average accessibility scores for all websites from each sector in reference to Table 3 in Figure 2. Here, the y-axis refers to the accessibility score. The sector-wide scores in this figure follow a similar trend as in Figure 1. Therefore, the accessibility scores for e-commerce (∼4.33), news (∼3.91), buy & sell (∼4.08), and job/freelance (∼ 4.08) websites are comparatively higher than

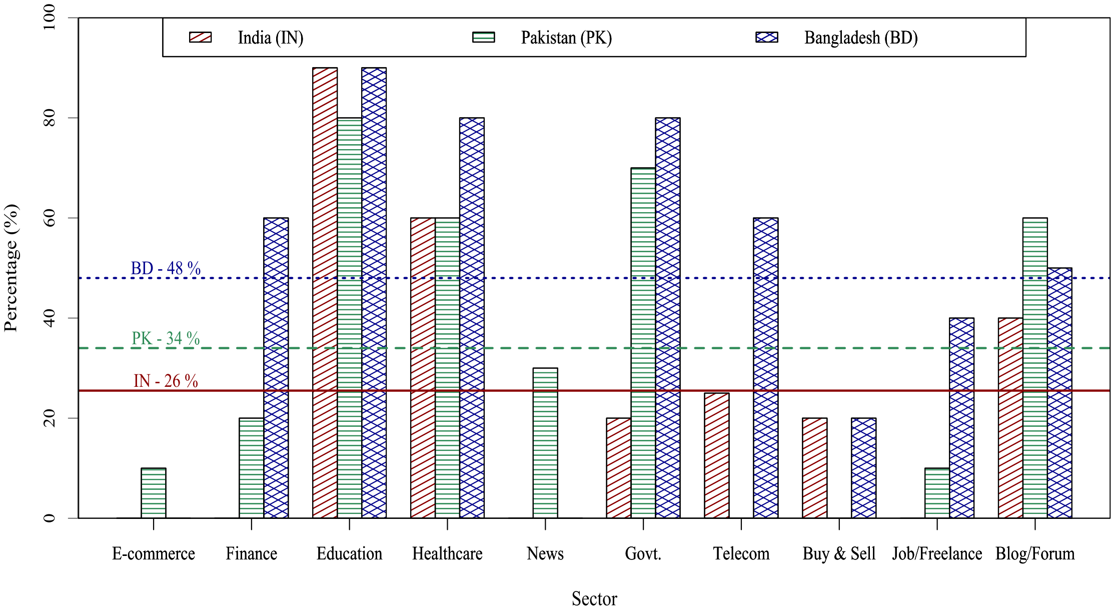


FIGURE 1. Percentage of websites without a privacy policy across all sectors.

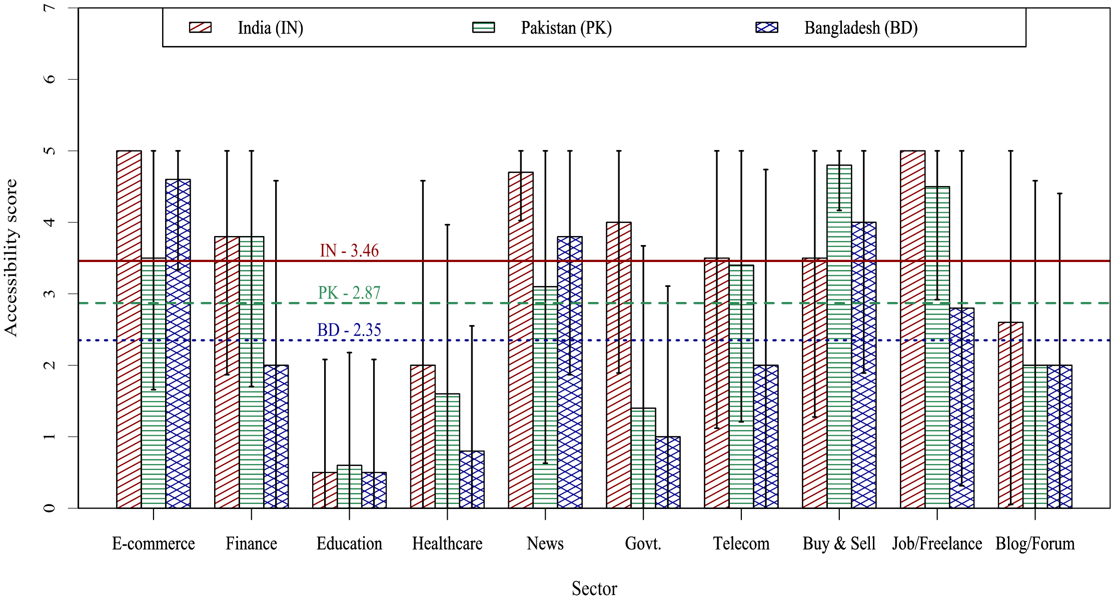


FIGURE 2. Average accessibility scores for website privacy policies across all sectors.

those in other sectors. Overall, websites in India, Pakistan, and Bangladesh across all sectors have an average scores of 3.46 ± 2.23, 2.87 ± 2.36, and 2.35 ± 2.43, respectively, as the horizontal lines show in the figure. Also, the large standard deviations in these scores correspond to the high variations in the accessibility scores for each sector, as illustrated by the whiskers of each bar in Figure 2. These scores imply that the websites in South Asia are not highly accessible to their users. In our dataset, only 152 out of the 284 websites (i.e., $\sim$54%) contain a clearly visible privacy statement on their landing pages.

B. READABILITY SCORES

Figure 3 presents a summary of average readability scores for the websites from each sector that have a privacy statement, i.e., they have an accessibility score $>$ 0. The scores represent Flesch-Kincaid grade levels,[1] as discussed in Section IV-C, which are plotted on the y-axis of this

[1] We computed Flesch-Kindcaid grade-level score only if the privacy statement of a website is available in English. For example, Figure 3 does not have a score for the Bangladeshi websites from the government sector because either they do not have privacy statements or they are written in the local language.

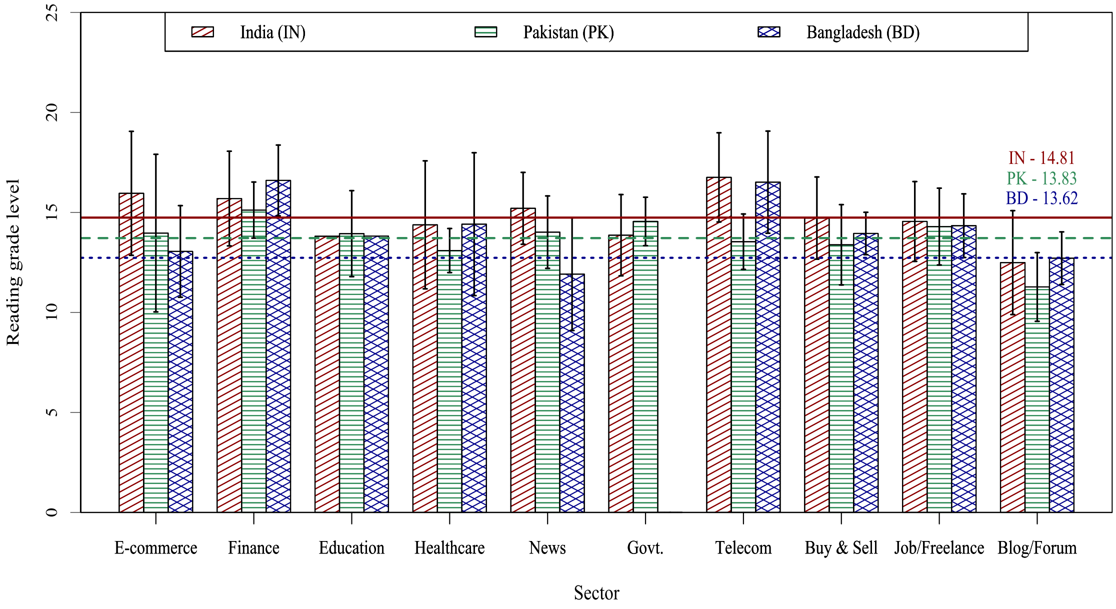


FIGURE 3. Average reading grade level of websites that have a privacy policy.

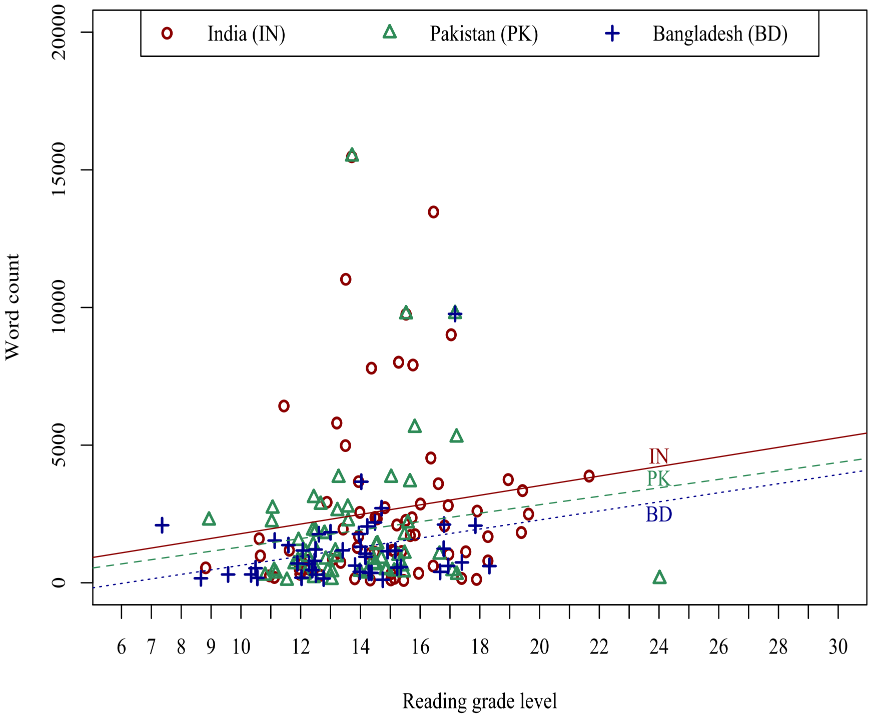
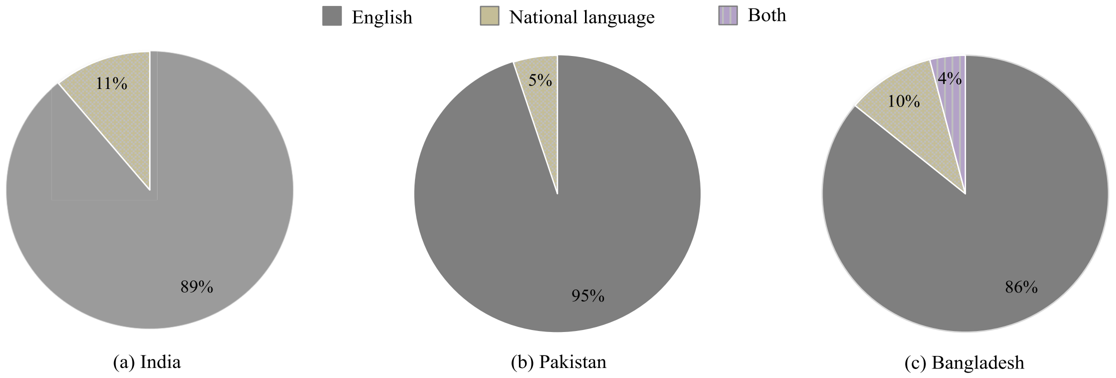


FIGURE 4. Reading grade level vs. word count of website privacy policies.

figure. This figure shows that the average readability score of the websites in all sectors is consistently high except the blog/forum sector which has the lowest score of 12.4. According to the Flesch-Kincaid grade-level formula, any score between 12+ to 15 corresponds to texts that are difficult to read since they are suitable for college-level readers [43]. This categorization is also corroborated by the country-wide average scores, as identified by the horizontal lines in Figure 3. For example, websites in India, Pakistan, and Bangladesh require users to have at least 14.81 ± 2.25, 13.83 ± 2.23, and 13.62 ± 2.43 years of formal education, respectively, to understand their privacy statements. Even though these scores are on par with popular websites, e.g., Yahoo, Facebook, and New York Times, their jargon/complexity



(a) India (b) Pakistan (c) Bangladesh

FIGURE 5. Language distributions of privacy policy for websites that have privacy statements.

most likely discourages users from reading their privacy policies [44].

Figure 4 plots the readability scores (x-axis) against the length (y-axis) of privacy statements to summarize their correlation in our dataset. A regression analysis of the country-wide data points is shown as the solid, dashed, and dotted diagonal lines for India, Pakistan, Bangladesh, respectively. The positive slopes of these lines infer that the reading complexity of privacy statements in South Asia increases linearly with their word counts. It is to be noted that shorter privacy statements do not necessarily guarantee an improved comprehension of privacy policies [15]; however, longer privacy statements are not desired either, as discussed in Section IV-C. The average word counts of the websites based in India, Pakistan, and Bangladesh are 2628, 1889 and 1232 words, respectively, as shown in Figure 4. Considering a typical college-level reading speed of ∼300 words per minute [45], users in these countries require 9, 6.5, and 4 minutes, respectively, to skim through the privacy statements. These reading costs may appear highly feasible; however, the high readability scores of the privacy statements should be a consideration in the above assessment. For example, the privacy statement of BBC website is suitable for readers with middle-school education and yet it requires 15 minutes of reading time [44].

The language distributions of the privacy statements for websites in India, Pakistan, and Bangladesh are shown in Figure 5. The figure illustrates the percentage of websites with an accessibility score > 0 that contain privacy statements in English only, national language[2] only, or both English and national language. These results show that South Asian websites put insignificant efforts in presenting privacy policies to their local users. For instance, only ∼6% websites have their privacy statements written in local languages. This finding is undesirable, for the literacy rate in this region is known to be low (Pakistan: 59%) or moderate (India and Bangladesh: 74%) [46].

C. COMPLIANCE SCORES

In this section, we discuss the extent to which the South Asian websites comply with privacy principles. Overall, we found that the average compliance score of the 284 South Asian websites in our dataset is low, i.e., 4.55. This means that on average, the websites comply with ∼4 out of 11 privacy principles, which are listed in Table 4. Figure 6 further shows that India has the highest overall compliance score of 5.97 ± 4.07, followed by Pakistan (4.66 ± 4.02) and Bangladesh (3.28 ± 3.81). The high variations in the aforementioned scores suggest that South Asian websites are not consistent enough in implementing data privacy across different sectors. For example, this figure shows that e-commerce and job/freelance websites scores are much higher than those in other sectors, e.g., education, healthcare, and government.

We present an analysis of privacy compliance[3] in reference to country-wide and sector-wide statistics in Tables 5 and 6, respectively. We discuss these statistics for each privacy principle below.

1) DATA PROCESSING (*COLLECTION*)

We found that a majority of the websites clearly outlined their data collection practice in the privacy statement. However, Table 5 suggests that only ∼57% of them fully comply with the *collection* part of the data-processing principle. It means that only a subset of these websites clearly state the personal information they collect from users, which primarily includes name, email, IP address and cookie information, across all sectors. Our country-wide analysis in the same table also shows that India has the highest percentage (∼73%) of websites that comply with this principle followed by Pakistan (∼56%) and Bangladesh (∼43%) respectively. Note that considering data processing, and more specifically, data collection, is the sole privacy principle outlined in DSA, the compliance score for Bangladesh is significantly low.

In terms of our sector-wide analysis, Table 6 suggests that e-commerce, buy & sell, and Job/freelance websites have

[2] Hindi, Urdu, and Bangla are the national/popular regional languages in India, Pakistan, and Bangladesh respectively.

[3] A more detailed compliance score analysis of South Asian websites is available in Appendix B.

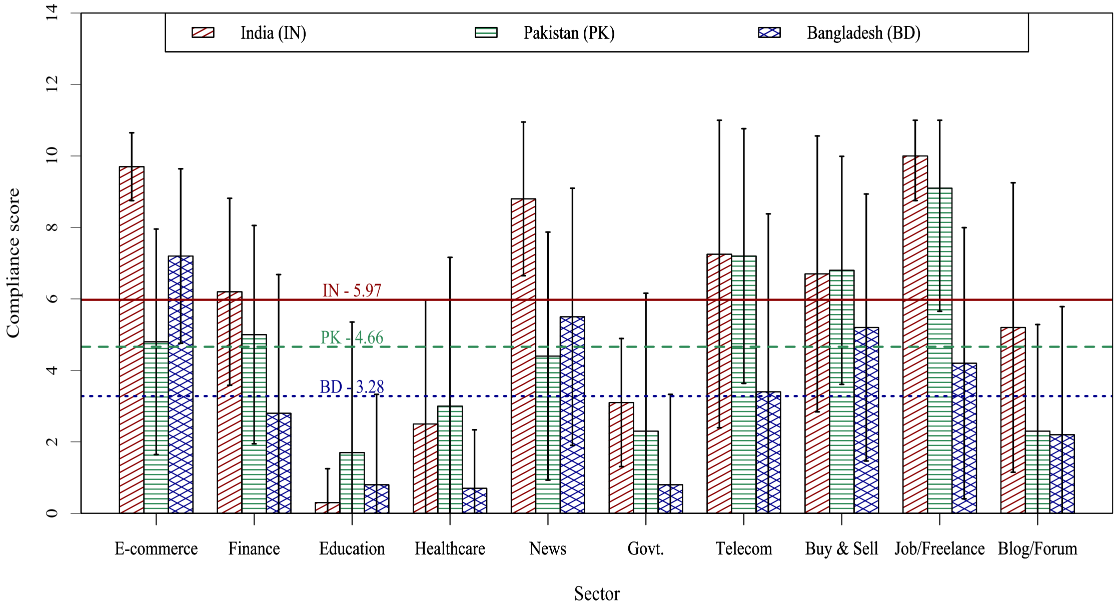


FIGURE 6. Average compliance scores of website privacy policies across all sectors.

TABLE 5. Overall percentage (%) of South Asian websites complying with the privacy principles.

	Privacy Principles										
	Data processing *(collection)*	**Data processing** *(purpose)*	**Protection of childern's data**	**Data accuracy & control**	**Data retention**	**Integrity & confidentiality**	**Accountability**	**Transparency**	**Data portability**	**Right to object**	**Third-party transfer**
Overall	57.04	59.85	18.30	43.30	25.35	52.46	46.12	45.77	27.11	45.42	58.45
India	73.40	76.59	26.59	62.76	41.48	69.14	62.76	65.95	37.23	65.95	77.65
Pakistan	56.84	61.05	21.05	41.05	26.31	58.94	52.63	50.52	35.78	48.42	69.47
Bangladesh	43.15	46.31	13.68	34.73	18.94	42.10	37.89	37.89	27.36	43.15	51.57

relatively high compliance, i.e., ∼80%, with data processing principle. This phenomenon can be attributed to the popularity, user volume, or financial interactions involved in these websites. However, the websites in other critical sectors such as healthcare, finance, and telecom have low compliance scores, i.e., 30%, ∼53%, and ∼64%, respectively. This is not unexpected since the websites in these sectors are static in design and are primarily used for informational purposes. For example, most South Asian banks and hospitals in our dataset do not provide online-banking and online-appointment services, respectively. Note that the websites from the education sector have the lowest compliance score of ∼13% and this score can be justified with the same argument discussed above.

2) DATA PROCESSING (*PURPOSE*)

This principle has the highest overall (∼59%) compliance in our dataset. Most of the websites either partially or fully state that the collected data will be used for notifying new services/products, providing better user experience, user analytics, or market research etc. The country-wide and sector-wide statistics for this part of the data processing principle are mostly in agreement with those in the data collection part. For example, India (∼76%) has the highest compliance, followed by Pakistan (∼61%), and Bangladesh (∼46%). Once again, the compliance of e-commerce, buy & sell, and job/freelance websites are high, whereas education websites seem to have the the lowest score among all sectors.

3) PROTECTION OF CHILDREN's DATA

This principle has the lowest overall compliance (∼18%) in our dataset. Even though IPDPB mandates the protection of children's data, we found that only ∼27% of the Indian websites comply with this principle. On the other hand, PPDPB and DSA lack in the protection of children's data, but around 21 and 14% websites in Pakistan and Bangladesh,

TABLE 6. Overall percentage (%) of South Asian websites per sector complying with the privacy principles.

Website Sector	Privacy Principles										
	Data processing (*collection*)	Data processing (*purpose*)	Protection of childern's data	Data accuracy & control	Data retention	Integrity & confidentiality	Accountability	Transparency	Data portability	Right to object	Third-party transfer
E-commerce	80	86.66	33.33	80	43.33	76.66	76.66	50	46.66	60	90
Finance	53.33	70	0	26.66	23.33	63.33	56.66	56.66	13.33	40	63.33
Education	13.33	13.33	3.33	13.33	3.33	6.66	10	6.66	3.33	10	10
Healthcare	30	33.33	6.66	10	6.66	26.66	10	20	6.66	30	26.66
News	76.66	76.66	20	53.33	36.66	66.66	53.33	60	43.33	63.33	73.33
Government	40	40	3.33	26.66	6.66	20	13.33	10	6.66	16.66	23.33
Telecom	64.28	64.28	14.28	42.85	50	71.42	50	57.14	50	50	71.42
Buy & Sell	83.33	80	26.66	56.66	23.33	73.33	60	70	26.66	50	73.33
Job/Freelance	83.33	83.33	53.33	76.66	46.66	80	70	80	50	76.66	76.66
Blog/Forum	46.66	46.66	10	33.33	10	30	40	26.66	6.66	26.66	46.66

respectively, still comply with this principle. Across all sectors, job/freelance has the highest compliance score (∼53%), whereas finance, education, government, and healthcare have significantly low scores ≤10%. It is to be noted that the e-commerce websites also fare significantly poorly (*sim*33%) here when compared with the two data-processing principles above.

4) DATA ACCURACY AND CONTROL

While analyzing this principle, we found that most of the websites did not specify all the required aspects of data accuracy and control, i.e., access, modification, and deletion of user data. Overall, ∼43 of these websites partially implement the above-mentioned aspects of this principle. In India and Pakistan, around 63% and 41% websites, respectively, comply with this principle. Once again, compliance with this principle is the lowest (∼34%) for Bengali websites. In terms of sector-wide analysis, e-commerce and job/freelance websites have the highest compliance (above 75%) whereas healthcare and education have the lowest compliance (below 14%) with this principle.

5) DATA RETENTION

We observed that many websites specify in their privacy statements that they retain the data collected from the users. However, only a small subset (∼25%) clearly mention about the duration of data storage or whether the data storage is limited to specific data-processing purposes. For this privacy principle, the country-wide compliance is low as well since none of the South Asian countries have a compliance score above 42%. Among the sectors, telecom websites (50%) are the most compliant with this principle followed by job/freelance and e-commerce, both scoring at ∼45%. As above, the least compliant (∼7%) websites are in education, healthcare, and government sectors.

6) INTEGRITY AND CONFIDENTIALITY

Our analysis shows that ∼52% websites in the dataset clearly mention about the implementation of various security measures (e.g., state-of-the art safeguards and encryption protocols) to ensure compliance with this principle. Here, ∼69% of Indian websites, ∼59% of Pakistani and ∼42% of Bangladeshi websites comply with this principle. Looking at the sectors, we find that e-commerce, buy & sell, and telecom websites have the highest compliance at ∼74% whereas the lowest compliance (∼7%) is found in education websites.

7) ACCOUNTABILITY

On average, less than half (e.g., ∼46%) of the websites in our dataset comply with the accountability principle. The compliance percentage for the websites in India, Pakistan, and Bangladesh is around 63%, 53%, and 38% respectively. The sector-wide data analysis repeats similar trend as above for the high (∼77) and low (10%) scores involving e-commerce and education/healthcare websites, respectively.

8) TRANSPARENCY

Similar to the previous principle, less than half of the websites (∼46%) notify the users (via email) about updates to privacy statements, or encourage them to periodically read their privacy statements for updates, which are sometimes labeled with a time stamp at the top/bottom of the statements. Our analysis across the countries shows that this principle is implemented in ∼66% Indian websites, followed by ∼51% Pakistani websites and ∼38% Bangladeshi websites. In terms of sector-wide analysis, we observe that the highest (80%) and the lowest (∼7%) compliance is in the job/freelance and education websites, respectively.

9) DATA PORTABILITY

The compliance for data portability is one of the lowest in our analysis. For example, ∼27% of all the websites state whether users can request a copy of their collected personal information, possibly, in a structured format. Among the three countries, India and Pakistan's compliance is ∼10% higher whereas Bangladesh fares the same percentage as mentioned above. Among the sectors, relatively high compliance at 50% is available only in telecom and job/freelance websites. Unfortunately, very low compliance (e.g., below 7%) is observed in websites from five different sectors: finance, education, healthcare, government, and blog/forum.

10) RIGHT TO OBJECT

Around 45% websites provide either an email address or other contact information to allow users to inquire about their privacy issues with a grievance officer. This trend is about the same among Pakistani and Bangladeshi websites whereas ∼66% websites in India provide the above information. Note that the lowest (10%) and the highest (∼77%) compliance scores for this principle are found in jobs/freelance and education websites, respectively.

11) THIRD-PARTY TRANSFER

The overall compliance for this principle is on par with that of the data-processing principles, i.e., high. ∼59% websites that we investigated clearly outlined how they share user data with third-party entities. This observation holds consistently in country-wide analysis because India, Pakistan, Bangladesh's scores are around 78, 69, and 52%, respectively. In sector-wide analysis, consistent high scores above 63% are available in majority of the sectors including e-commerce, telecom, and news. Low scores below 30% are found in education and healthcare websites as observed earlier.

VI. DISCUSSION

The intended contribution of this study is to perform a qualitative assessment that can be leveraged to improve privacy practices in South Asia. Our stakeholders are website owners, users, and privacy regulators. We provide practical implications in these aspects along with a comparative summary of our results.

A. SUMMARY OF FINDINGS

1) ACCESSIBILITY

It appears that the accessibility of privacy statements in South Asian websites is low but this finding is not surprising if we compare it with existing work. For example, websites in government [16], [30] and healthcare [26], [31] sectors usually have low accessibility scores around the globe. However, this is not true for e-commerce [13] and finance [15] websites. This phenomena demands a greater degree of awareness among website owners and users for improving this privacy aspect in South Asia.

2) READABILITY

The readability scores of the investigated websites are roughly on par with popular websites around the globe, e.g., Facebook and New York Times, concerning their Flesch-Kincaid reading grade levels. However, this aspect needs to be improved significantly for the region's poor literacy rate and pervasive usage of local languages, as discussed in Section V-B. Here, the website owners have to take preemptive measures for providing privacy statements in a manner that is clear and transparent for South Asian users.

3) COMPLIANCE

The compliance scores of our websites are relatively low, with only 23 out of 284 websites, mostly in e-commerce, fully complying with all 11 privacy principles. None of the investigated banking websites complied with all 11 principles. Such low privacy compliance is found to be consistent with other studies conducted on finance [15], healthcare [31], and government [30] sectors. South Asian websites appear to be less likely to protect children's data and implement data retention and portability policies. This observation is partially in agreement with a prior study on mobile money and banking services that showed low compliance with children's data protection, and accountability principles. Even though no study concerning all sectors investigated here is available, the compliance of South Asian websites with essential privacy principles is low, e.g., ∼47%, in general. Hence, all stakeholders have to come together to improve, enforce, and practice privacy principles in this region.

B. PRACTICAL IMPLICATIONS

There are many practical implications of this study. Our analysis identified areas of improvement after studying a large number of websites based in India, Pakistan, and Bangladesh. This can assist various stakeholders involved in the implementation of user privacy in the respective countries.

1) WEBSITE OWNERS

Website owners can improve the accessibility, readability, and compliance of their privacy statements by incorporating the following suggestions:

- Provide a link to privacy statements on the landing page to improve accessibility.
- Provide easily comprehensible privacy statements with a reading grade level of 8 or less.
- Provide privacy statements in native/regional languages of South Asia.
- Improve compliance by implementing all privacy principles (specified by the regional data protection regulation). For instance, give protection of children data, data retention, data portability equal importance as other principles.
- Provide contact information of grievance officer in the privacy statements to enable users to exercise their right to object a principle or to know more about their privacy.

TABLE 7. Website dataset.

Sector	India	Pakistan	Bangladesh
E-commerce	Amazon India Flipkart Alibaba Snapdeal Myntra IndiaMART Book My Show Nykaa First Cry 1mg	Daraz.pk Juniba.pk Chase value centre Telemart.pk Yayvo.com The Warehouse.pk Vmart.pk Shophive Shopdaily.pk Homeshopping.pk	Rokomari Daraz Ajkerdeal Pickaboo Bagdoom Othoba Priyoshop Banglashoppers iferi Chaldal
Finance	State Bank of India ICICI Bank HDFC Bank Axis Bank Kotak Mahindra Bank IndusInd Bank Bank of Baroda Panjab National Bank YES Bank IDBI Bank	Habib Bank Limited National Bank of Pakistan Meezan Bank MCB United Bank Limited Allied Bank Standard Chartered Bank Bank Alfalah Askari Bank Faysal Bank	HSBC Dutch-Bangla Bank Sonali Bank Limited Islami Bank Limited Grameen Bank Janata Bank Limited Standard Chartered Bank Prime Bank Limited Habib Bank Limited State Bank of India
Education	IIT Bombay IIT Madras IIT Kanpur IIT Delhi IIT Kharagpur University of Delhi Tata Inst. of Fundamental Research Indian Inst. of Science Banglore Indian Inst. of Tech. Roorkee Vellore Inst. of Tech.	Quaid-i-azam University COMSATS University University of Agriculture International Islamic Uni. University of Karachi National Uni. of Sci. & Tech. University of the Punjab Bahauddin Zakariya Uni. Aga Khan University Government College Uni.	BUET University of Dhaka Rajshahi University SUST BRAC University Jahangirnagar University North South University University of Chittagong Independent University KUET
Healthcare	Aravind Eye Care System Laparoscopy Hospital Tata Memorial Centre Max Healthcare Sri Ramachandra Uni. & Medical Center Amrita Inst. of Med. Sci. & Research Center Sri Venkateswara Inst. of Med. Sci. Sankara Nethralaya Hospital M.V. Hospital for Diabetes Mahatma Gandhi Inst. of Med. Sci.	Indus Hospital Services Inst. of Med. Sciences Sindh Institute of Urology & Transplantation Shifa International Hospital Shaukat Khanum Memorial Cancer Hospital Al Shifa Trust Eye Hospital Pak. Inst. of Med. Sciences LRBT Eye Hospital Liaquat National Hospital Allama Iqbal Med. College & Jinnah Hospital	Center for the Rehabilitation of the Paralysed Jalalabad Ragib-Rabeya Med. College Mymensingh Medical College Ahsania Mission Cancer & General Hospital Armed Forces Med. College Hospital United Hospital Limited Marie Stopes Health Clinic Square Hospital Limited Northern Intl. Med. College Hospital Ibn Sina Trust
News	The Times of India Malayala Manorama NDTV India Today One India Hindustan Times Mid Day India TV ABP News Parda Phash	Geo News ARY News Dunya News Samaa News Tribune 92 News Dawn News Aaj News Express News Bol News	Prothom Alo Jugantor Kalerkantho Bandgladesh Protidin Jagonews24 Somoi News TV Bdnews24 Banglanews24 Ittefaq The Daily Star

2) PRIVACY REGULATORS

Our findings emphasize the need for privacy regulators in Pakistan and Bangladesh to improve their existing laws by incorporating the following changes:

- Include information on protection of children's data in PPDPB and DSA.
- Extend DSA beyond data processing, or legislate a new privacy law in Bangladesh in reference to GDPR.

3) END USERS

Our findings can help raise awareness among website users in South Asia. Users do not read privacy statements not only because they are long but also because of the assumption and inherent trust in the website, that it is ensuring proper mechanisms to protect their personal data. Our findings serve as a summary of the privacy statements for the users and suggest that the users in South Asia be cautious when visiting websites especially in the healthcare, government, and education sectors. Users should also be aware that the current laws in Pakistan and Bangladesh do not specify protections for children's data. Our findings also caution the users into looking for the contact information of grievance officer when visiting a website that requires sensitive information such as credit card number. This is so that the users can inquire about their privacy rights, amend their information, or request a copy of their data collected by the respective website.

C. LIMITATIONS AND FUTURE WORK

Our study is not without limitations; therefore, we see various opportunities for future research. First, we performed a qualitative analysis on a dataset of 284 websites by including top

TABLE 8. Website dataset.

Sector	India	Pakistan	Bangladesh
Government	Ministry of Road Transport and Highways Department of Commerce Planning Commission Ministry of Labour and Employment Ministry of Home Affairs Ministry of External Affairs Ministry of Human Resource Development Ministry of Defense Ministry of Info. & Broadcasting Ministry of Finance	NADRA Federal Board of Revenue Ministry of Defense Ministry of Foreign Affairs Ministry of Education Overseas Pakistanis Foundation Pak. Export Processing Zones Authority Ministry of Info. Tech. & Broadcasting Overseas Employment Corp. National Highway Authority	BD Road Transport Authority BD Export Processing Zones Authority BD Planning Commission Ministry of Expatriates' Welfare & Overseas Employment Ministry of Home Affair Ministry of Foreign Affairs Ministry of Education Ministry of Defense Ministry of Information National Board of Revenue
Telecom	Jio Vodafone Idea Airtel BSNL	Ufone Warid/Jazz PTCL Zong Telenor	Grameenphone Banglalink Robi Axiata Teletalk BTCL
Buy & Sell	Limeroad Groupon CraftGhar Saree Rediff Shopping Eleb2b CraftsVilla eBay OLX gocoop	OLX Asan Classifieds bolee Go Cheap Shop Zameen.com PakWheels.com CoinBolee buyon.pk Dealmarkaz aarz	Clickbd cellbazaar.com bikroy.com BD Bazar24 FastBikri BuySellBazar24.com Trade Bangla BizBangladesh.com bdhousing.com komdaame
Job/Freelance	Naukri Shine Monster Freshersworld LinkedIn FreelanceMyWay Indeed Glassdoor JobSarkari Upwork	Rozee.pk Mustakbil.com Jobee.pk Jobz.pk Indeed 99Designs People Per Hour Freelancer Upwork Fiverr	bdjobs.com Skill.jobs Youth Opportunities bdjobstoday.com ShadhinKaj OutsourceMyJob KajKey Freelancer Bangladesh Spotlight Bangladesh Careerjet
Blog/Forum	Digital Bhoomi Fropky India Forum India Forum Discussion Reddit India Community Team BHP-Forum Indian Consumer Complaints 99 Acers IndiaVideoGamer MoneyControl	Urdupoint Hamariweb Propakistani.pk Janubaba.com Siasat.pk itdunya cssforum.com.pk Pakistanipoint.com Gupshup.org Friendskorner.com	Sachalayaton Cadet College Blog somewhereinblog.net Projonmo Forum TechTunes Choturmatrick BanglaCricket.com Tarunyo mukto-mona.com Bangla Hub

10 websites from 10 sectors. Expanding the current dataset by increasing the number of observations in each country to perform a more detailed statistical analysis is a potential avenue for future work. Secondly, our analysis was manual. The reason behind this manual analysis stems from the fact that privacy statements are written in a natural language and their structures are usually not uniform. An extension of this work will look into developing a browser extension based on the proposed methodology for computing compliance scores for a visited website on the fly. Lastly, we were unable to analyze the readability and privacy compliance of websites whose privacy statements are only available in a regional language due to the poor accuracy of publicly available translation services, e.g., Google Translate. We are considering a future work that will study the machine translation of privacy statements to determine the accuracy of available translation services.

VII. CONCLUSION

Privacy is a new concept among website users in South Asia. However, awareness of this concept is growing in this region as is evident from the recent efforts of privacy legislations based on GDPR in India, Pakistan, and Bangladesh. In this study, we performed a qualitative assessment of privacy statements of 284 popular websites across 10 different sectors in the above-mentioned countries. We analyzed the availability and readability scores of privacy statements for these websites using a scoring rubric and Flesch-Kincaid formula, respectively. We also used a keyword-based content analysis to evaluate the compliance of 11 GDPR privacy principles by the websites. This is an original work since no study on South Asian websites are available to the best of our knowledge.

Our results show that privacy statements of South Asian websites are not easily accessible to users as they are mostly available on Frequently Asked Questions or Terms & Conditions pages. The readability of the statements are not suitable for users since they require college-level education for reasonable comprehension of the policy contents. Most importantly, only 23 websites in our dataset fully comply with all 11 GDPR principles. In brief, websites in India, Pakistan, and Bangladesh implement around 6, 5 and 3 GDPR

TABLE 9. Percentage (%) of Indian (IN), Pakistani (PK), and Bangladeshi (BD) websites complying with privacy principles.

Website Sector	Country	Privacy Principles										
		Data processing (*collection*)	Data processing (*purpose*)	Protection of childern's data	Data accuracy & control	Data retention	Integrity & confidentiality	Accountability	Transparency	Data portability	Right to object	Third-party transfer
E-commerce	IN	100	100	70	100	70	100	100	100	40	90	100
	PK	50	60	10	60	20	60	60	10	40	30	80
	BD	90	100	20	80	40	70	70	40	60	60	90
Finance	IN	70	90	0	30	30	100	90	70	10	50	80
	PK	70	80	0	20	10	60	50	80	10	40	80
	BD	20	40	0	30	30	30	30	20	20	30	30
Education	IN	10	10	0	10	0	0	0	0	0	0	0
	PK	20	20	0	20	10	20	20	10	10	20	20
	BD	10	10	10	10	0	0	10	10	0	10	10
Healthcare	IN	40	40	0	10	10	40	10	20	10	30	40
	PK	40	40	20	20	10	30	20	40	10	40	30
	BD	10	20	0	0	0	10	0	0	0	20	10
News	IN	100	100	30	70	70	80	80	100	60	100	90
	PK	60	70	20	30	0	60	30	40	30	40	60
	BD	70	60	10	60	40	60	50	40	40	50	70
Government	IN	80	80	0	60	0	20	0	0	0	40	30
	PK	30	30	10	10	20	30	30	20	10	10	30
	BD	10	10	0	10	0	10	10	10	10	0	10
Telecom	IN	75	75	0	75	75	75	50	75	75	75	75
	PK	80	80	20	40	60	100	60	60	60	60	100
	BD	40	40	20	20	20	40	40	40	20	20	40
Buy & Sell	IN	80	80	30	70	30	70	70	80	30	50	80
	PK	90	90	20	70	30	80	70	70	40	40	80
	BD	80	70	30	30	10	70	40	60	10	60	60
Job/Freelance	IN	100	100	70	100	80	100	90	100	60	100	100
	PK	90	90	80	80	60	90	90	90	70	90	80
	BD	60	60	10	50	0	50	30	50	20	40	50
Blog/Forum	IN	70	70	20	70	20	50	60	40	20	30	70
	PK	40	40	0	20	10	20	30	10	0	20	40
	BD	30	30	10	10	0	20	30	30	0	30	30

principles, respectively. Our sector-wide analysis suggests that e-commerce websites are most compliant and all sectors clearly outline data collection and sharing practices with their users. It is, however, notable that all websites make the least effort in protecting children's data even though it is an important privacy principle in GDPR. Based on these results, we provide recommendations involving all stakeholders (i.e., website owners, privacy regulators, and users) to help improve privacy protection of user data in South Asia.

APPENDIXES

APPENDIX A

WEBSITE DATASET

Tables 7 and 8 show the names of the 284 websites from 10 different sectors in India, Pakistan, and Bangladesh that we have used in our study. Our datasets along with the websites' homepage and privacy-policy page URLs are also publicly available on IEEE *DataPort* (DOI:10.21227/fkap-re05).

APPENDIX B

DETAILED COMPLIANCE SCORES

Table 9 shows a detailed summary of the compliance scores for the sector-wide websites in India, Pakistan, and Bangladesh. This table presents individual scores for each privacy principle that we have used to evaluate South Asian websites.

REFERENCES

[1] D. J. Solove, "Conceptualizing privacy," *California Law Rev.*, vol. 90, no. 4, p. 1087, Jul. 2002.

[2] D. Solove, *Understanding Privacy*. Cambridge, MA, USA: Harvard Univ. Press, 2008.

[3] F. Cate, *Privacy in the Information Age*. Washington, DC, USA: Brookings Institution Press, 1997.

[4] E. F. Stone, H. G. Gueutal, D. G. Gardner, and S. McClure, "A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations.," *J. Appl. Psychol.*, vol. 68, no. 3, pp. 459–468, 1983.

[5] H. Smith, S. Milberg, and S. Burke, "Information privacy: Measuring individuals' concerns about organizational practices," *MIS Quart.*, vol. 20, no. 2, pp. 167–196, 1996.

[6] *General Data Protection Regulation (GDPR)–Official Legal Text*. Accessed: Jul. 25, 2020. [Online]. Available: https://gdpr-info.eu/

[7] H. Perera, W. Hussain, D. Mougouei, R. A. Shams, A. Nurwidyantoro, and J. Whittle, "Towards integrating human values into software: Mapping principles and rights of GDPR to values," in *Proc. IEEE 27th Int. Requirements Eng. Conf. (RE)*, Sep. 2019, pp. 404–409.

[8] F. Cate. (2006). *The Failure of Fair Information Practice Principles*. Consum. protection age Inf. economy. [Online]. Available: https://ssrn.com/abstract=1156972

[9] A. Satariano. (Jan. 21, 2019). *Google is Fined $57 Million Under Europe's Data Privacy Law*. The New York Times, Accessed: Jul. 25, 2020. [Online]. Available: https://www.nytimes.com/2019/01/21/technology/google-europe-gdpr-fine.html

[10] (2020). *Data for South Asia, India, Pakistan, Bangladesh, Afghanistan, Bhutan, Maldives, Nepal, Sri Lanka*. Accessed: Jul. 25, 2020. [Online]. Available: https://data.worldbank.org/?locations=8S-IN-PK-BD-AF-BT-MV-NP-LK

[11] (Jan. 1, 1970). *8 Biggest Data Leaks Of 2019 That Hit Indian Users Hard–What Causes Data Breach?*. The India Times. Accessed: Jul. 25, 2020. [Online]. Available: https://economictimes.indiatimes.com/industry/tech/8-biggest-data-leaks-of-2019-that-hit-indian-users-hard/what-causes-data-breach/slideshow/72839190.cms

[12] A. Mishbah. (Apr. 7, 2019). *Bangladesh steps into the Data Protection Regime*. The Daily Star. Accessed: Jul. 25, 2020. [Online]. Available: https://www.thedailystar.net/opinion/human-rights/news/bangladesh-steps-the-data-protection-regime-1726351

[13] M. Desai, K. Desai, and L. Phelps, "E-commerce policies and customer privacy: A longitudinal study (2000–2010)," *Inf. Manage. Comput. Secur.*, vol. 20, no. 3, pp. 222–244, 2012.

[14] R. Zaeem and K. Barber, "A study of Web privacy policies across industries," *J. Inf. Privacy Secur.*, vol. 13, no. 4, pp. 169–185, 2017.

[15] J. Bowers, B. Reaves, I. N. Sherman, P. Traynor, and K. Butler, "Regulators, mount up! analysis of privacy policies for mobile money services," in *Proc. 13th Symp. Usable Privacy Secur.*, San Diego, CA, USA, 2017, pp. 97–114.

[16] S. M. Alhomod and M. M. Shafi, "Privacy policy in e government Websites: A case study of Saudi Arabia," *Comput. Inf. Sci.*, vol. 5, no. 2, p. 88, Feb. 2012.

[17] (2002). *Privacy Policies*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.privacypolicies.com/blog/privacy-policies-legally-required/

[18] (2016). *European Data Protection Supervisor*. Accessed: Jul. 25, 2020. [Online]. Available: https://edps.europa.eu/data-protection_en

[19] (2019). *The Personal Data Protection Bill*. PRSIndia. Accessed: Jul. 25, 2020. [Online]. Available: https://www.prsindia.org/billtrack/personal-data-protection-bill-2019

[20] A. Burman and S. Rai. (Mar. 9, 2020). *What is in India's Sweeping Personal Data Protection Bill?*. Accessed: Jul. 25, 2020. [Online]. Available: https://carnegieindia.org/2020/03/09/what-is-in-india-s-sweeping-personal-data-protection-bill-pub-80985

[21] (2020). *Pakistan Data Protection Bill*. Ministry of Information Technology and Telecommunication. Accessed: Jul. 25, 2020. [Online]. Available: https://moitt.gov.pk/SiteImage/Misc/files/Personal%20Data%20Protection-%20Bill%202020%20Updated(1).pdf

[22] E. Foo and J. Tan. (May 2020). *Pakistan Releases Updated Draft Of Personal Data Protection Bill*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.twobirds.com/en/news/articles/2020/global/pakistan-releases-updated-draft-of-personal-data-protection-bill

[23] (Oct. 2018). *Bangladesh Digital Security Act*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.cirt.gov.bd/wp-content/uploads/2018/12/Digital-Security-Act-2018-English-version.
pdf

[24] (2018). *European economic area (EEA)/Relations with the EU*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.efta.int/eea

[25] G. Dias, H. Gomes, and A. Zúquete, "Privacy policies in Web sites of Portuguese municipalities: An empirical study," *Advances in Information Systems and Technologies*. Berlin, Germany: Springer, 2013, pp. 87–96.

[26] S. A. Rains and L. A. Bosch, "Privacy and health in the information age: A content analysis of health Web site privacy policy statements," *Health Commun.*, vol. 24, no. 5, pp. 435–446, Jul. 2009.

[27] C. Liu and K. P. Arnett, "Raising a red flag on global WWW privacy policies," *J. Comput. Inf. Syst.*, vol. 43, no. 1, pp. 117–127, 2002.

[28] (2001). *FDIC's Privacy Rule Handbook*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.fdic.gov/regulations/examinations/financialprivacy/handbook/-index.html

[29] (Jan. 1, 2016). *Mobile Privacy Principles. Promoting Consumer Privacy in the Mobile Ecosystem*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.gsma.com/publicpolicy/wp-content/uploads/2012/03/GSMA2016_Guidelines_Mobile_Privacy_Princi-ples.pdf

[30] A. D. Beldad, M. De Jong, and M. F. Steehouder, "When the bureaucrat promises to safeguard your online privacy: Dissecting the contents of privacy statements on dutch municipal Websites," *Government Inf. Quart.*, vol. 26, no. 4, pp. 559–566, Oct. 2009.

[31] J. Kuzma, K. Dobson, and A. Robinson, "An examination of privacy policies of global On-line E-pharmacies," *Eur. J. Res. Reflection Manage. Sci.*, vol. 4, no. 6, pp. 23–28. 2016.

[32] C. Robles-Estrada, J. Vargas-Barraza, and M. Sepúlveda-Nú nez, "Are privacy issues important in Mexican online markets? An empirical investigation into published online privacy statements of Mexican Web sites," in *Proc. BLED*, 2006, PP. 1–19. [Online]. Available: http://aisel.aisnet.org/bled2006/6

[33] *Southern Asia Population 2020 (Demographics, Maps, Graphs)*. Accessed: Jul. 25, 2020. [Online]. Available: https://worldpopulationreview.com/continents/southern-asia-population

[34] J. McCarthy. (Aug. 2, 2017). *Indian supreme court declares privacy a Fundamental Right*. Nat. Public Radio. Accessed: Accessed: Jul. 25, 2020. [Online]. Available: https://www.npr.org/sections/thetwo-way///2017/08/24/545963181/indian-supreme-court-declares-privacy-a-fundamental-right

[35] *Alexa–Keyword Research, Competitive Analysis, And Website Ranking*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.alexa.com/

[36] *Ranking Web of Universities, Cybermetrics Lab*. Accessed: Jul. 25, 2020. [Online]. Available: http://www.webometrics.info/en

[37] *Quora*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.quora.com/

[38] Z. K. Shalhoub, "Content analysis of Web privacy policies in the GCC countries," *Inf. Syst. Secur.*, vol. 15, no. 3, pp. 36–45, Jul. 2006.

[39] J. Kincaid, R. Fishburne, Jr., R. Rogers, and B. Chissom, "Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel," Nav. Tech. Training Command Millington TN Res. Branch, Univ. Central Florida, Tech. Rep. 8-75, 1975. Accessed: Feb. 1, 2017. [Online]. Available: https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=ist-library

[40] E. Beaunoyer, M. Arsenault, A. M. Lomanowska, and M. J. Guitton, "Understanding online health information: Evaluation, tools, and strategies," *Patient Edu. Counseling*, vol. 100, no. 2, pp. 183–189, Feb. 2017.

[41] M. Adamovic. (2009). *Readability Calculator*. Accessed: Jul. 25, 2020. [Online]. Available: https://www.online-utility.org/

[42] A. McDonald and L. Cranor, "The cost of reading privacy policies," *I/S: A J. Law Policy for Inf. Soc.*, vol. 4, pp. 543–568, 2008.

[43] R. Flesch, *How to Write Plain English: A Book for Lawyers and Consumers*. New York, NY, USA: Harper & Row, 1979.

[44] K. Litman-Navarro, "We read 150 privacy policies. They were an incomprehensible disaster," The New York Times. 2019. Accessed: Jul. 25, 2020. [Online]. Available: https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html

[45] R. P. Carver, "Silent reading rates in grade equivalents," *J. Reading Behav.*, vol. 21, no. 2, pp. 155–166, Jun. 1989.

[46] (2018). *Literacy Rate, Adult Total (% Of People Ages 15 And Above)*. UNESCO Institute for Statistics. Accessed: Jul. 25, 2020. [Online]. Available: https://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=8

YOUSRA JAVED received the B.S. and M.S. degrees in information and communication systems engineering from the National University of Sciences and Technology, Pakistan, in 2011, and the Ph.D. degree in computing and information systems from the University of North Carolina, Charlotte, NC, in 2017. From 2012 to 2014, she worked as a Research Assistant with the Laboratory of Information Integration Security and Privacy, Charlotte. After receiving the Ph.D. degree, she served as an Assistant Professor with the School of Information Technology, Illinois State University, from 2017 to 2019. She is currently an Assistant Professor with the School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Pakistan. Her research interests include usable security, privacy, and human computer interaction.

KHONDAKER MUSFAKUS SALEHIN (Senior Member, IEEE) received the B.S. degree in computer science and engineering from the Ahsanullah University of Science and Technology, Dhaka, Bangladesh, and the M.S. and Ph.D. degrees in telecommunications and electrical engineering from the New Jersey Institute of Technology (NJIT), Newark, NJ, USA. He is currently an Assistant Professor with the Department of Computer Science, California State University, Dominguez Hills, Carson, CA, USA. His research interest includes in computer networks and other related topics. He regularly serves as a TPC member or organizer for various IEEE conferences and as a reviewer or editor for different reputed journals. He was a recipient of multiple awards including Hashimoto Fellowship (2011-2013) and Outstanding Graduate Student Award (2013) from the Department of Electrical and Computer Engineering, NJIT, and New Jersey Inventors Hall of Fame-Graduate Student Award (2013).

MOHAMED SHEHAB received the Ph.D. degree in electrical and computer engineering from Purdue University, in August 2007. He is currently an Associate Professor with the Department of Software and Information Systems, University of North Carolina at Charlotte. He is the Director of the Laboratory of Information Integration, Security, and Privacy (LIISP). His research and teaching interests are in the broad areas of network and information security. In particular, his research focuses on advancing the state of the art in the design and implementation of distributed access control protocols to cope with the requirements of emerging distributed frameworks, mobile applications, web services, social networks, and database systems. He leads several research projects that focus on mobile application security, and mobile authentication. He is also interested in research topics related to security, privacy for third party applications and online social network applications, and usable security.