Afghanistan Research Journal – Natural Science

Volume 3, Issue 1, pp. 21-27, March 2022

ISSN (Online): 2789-8601

Privacy Risk in Recommender Systems

Khadija RAHMANY[1], Sayyed Kamran HOSSEINI[2]

1. Department of Software Engineering, Faculty of Computer Sceince, Herat University, Herat, Afghanistan
2. Department of Software Engineering, Faculty of Computer Sceince, Herat University, Herat, Afghanistan

Abstract

Nowadays, recommender systems are mostly used in many online applications to filter information and help users in selecting their relevant requirements. It avoids users to become overwhelmed with the massive amount of possible options. To provide an efficient and accurate personalized recommendation, such systems require a large amount of data of user's personal data which can provide by collecting privacy sensitive data from users such as ratings, consumption histories, and personal profiles. However, the privacy risks in gathering and processing personal data are often underestimated or ignored. The common privacy risks associated with recommender systems are the lack of adequate implementation of privacy protection principles. This review article aimed to evaluate the privacy risks in recommender systems. This paper discusses recommender systems and privacy concepts. Then, it gives an overview of the data that are used in recommender systems and examines the associated risks to data privacy. After that, the paper discusses relevant research areas for privacy-protection techniques and their applicability to recommender systems. The paper discussed various insights of user privacy, in both technical and non-technical environments, privacy design strategies, and privacy engineering approaches for developing a privacy-friendly recommender system. Finally, the paper concludes with a discussion on applying and combining different privacy-protection techniques. The results indicated that better user privacy can be achieved if privacy is considered by design and by default. Moreover, prediction accuracy is not limited by better user privacy when the privacy by architecture is considered alongside the privacy by design.

Keywords: Recommender systems, Privacy risk, Privacy design strategies.

INTRODUCTION

It is undeniable fact that information technology (IT) has come to every sector of human beings and people get benefit from its services in their lives. There are many systems that are developed to serve people in different areas to decide more suitably, efficiently, and faster. With social networks development, recommender systems have been increasingly prevalent and have become widely accepted by users. Recommended systems provide an automatic and personalized selection of data or items based on knowledge or data which are taken from these systems. Currently, most of the sites, over electronic media, use a recommender system to filter and customize the abundant amount of available choices according to the user's preferences and needs. Controlling information breaches in electronic environments, where information has a longer period of persistence, is highly matter in systems engineering. A common risk associated to privacy in recommender systems are methods of personalizing users' preferences according to gathered information, whether explicitly by asking questions or implicitly by tracking their behavioral activities.

However, the privacy risks in gathering and processing personal data are often underestimated or ignored. The goal of this article is incorporating a better user privacy in recommender systems without limiting their predictive results. This paper discusses privacy and privacy laws to build a privacy-friendly recommendation system.

RESEARCH METHODS

The paper methodology requires searching and gathering relevant information from textbooks. There are many books, journals and papers that talk about different aspects of recommender systems, user privacy risks, and frameworks. The newer publications are precedence over older ones as the author wants to show new researches. All these references show the importance of privacy risk in recommender systems. Overview of the researches helps to explore, analyze and collect the data which is necessary for the article. It takes many days to search which research depends on the topic. Most of the research were presented a recommender system and privacy concepts. The paper studies and compares different research papers and discusses the result in this paper. This review of the literature was performed in compliance with the PRISMA guidelines for systematic review. Data was collected from electronic databases such as Google scholar. Searching terms include "recommender systems", "user privacy risks", and "frameworks" connected by a Boolean operator "AND" (e.g., google scholar search strategy: "recommender systems" AND "user privacy risks"). Several textbooks were hand searched. All the titles as well as abstracts that appeared from this search, were reviewed. Specific inclusion criteria and exclusion criteria which were used to select the studies for review are mentioned in table 1.

Table 1. inclusion and exclusion criteria used to select studies for review

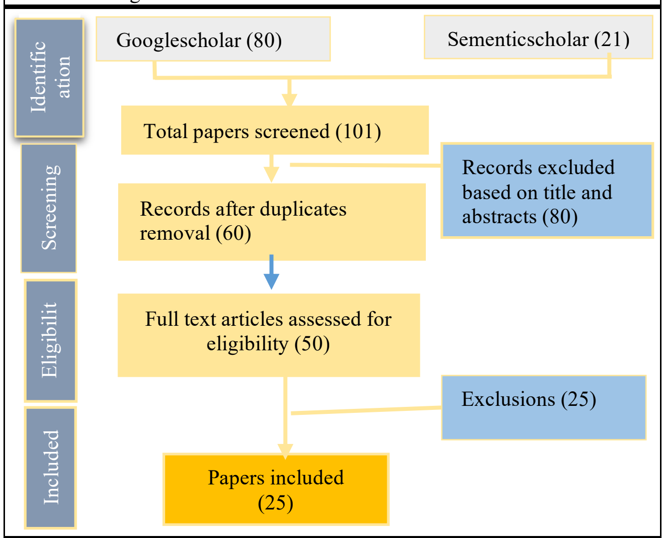
Inclusion criteria	Exclusion criteria
- Full-text available - Systematic review - Keyword findings - Studies were about Privacy in RS	- Narrative review - Unavailability of full-text article

When the title and abstracts were considered relevant the full-text of the article was reviewed. Full-text of eligible articles were also evaluated to ensure the article contents were relevant. All information and the outcomes were collected from article.

RESULTS

The search process resulted in total of 101 articles. After a preliminary screening of the titles and abstracts 80 irrelevant article and duplicates articles were excluded. The articles which their full-text was available (60 articles) were collected. Finally, after studying full-text of 25 articles were recorded according to the inclusion and exclusion criteria (Figure 1).

Figure 1. Flow chart of the search strategy and steps of this review based on PRISMA guidelines


Identific ation
Googlescholar (80)
Sementicscholar (21)
Total papers screened (101)
Screening
Records excluded based on title and abstracts (80)
Records after duplicates removal (60)
Eligibilit
Full text articles assessed for eligibility (50)
Exclusions (25)
Included
Papers included (25)

PRIVACY

Privacy always has been defined differently from different perspectives. As a real-world example with this regard can be referred to the discussion of legal privacy cultures between Europeans and Americans. Europeans respect personal dignity while Americans prefer freedom from intrusions, for example by state (Whitman, 2004).

Altman and Westin explained privacy as a "boundary" that how and to what extent information about a person can be communicated or circulated among others (Altman, 1975) (Westin, 1976). Nowadays, data integrity can be seen in Facebook providing groups for a specific group of people such as friends or family, in order to communicate with each other.

Today controlling information breaches in electronic environments, where information has a longer period of persistence, is important in systems engineering. Therefore, privacy should be considered as a set of definitions where each element (definition) has nothing in common but are the resemblance to each other (Solove, 2005). Generally, privacy discuses from two perspectives: non-technical perspective and technical perspective.

NON-TECHNICAL PERSPECTIVE

There are certain international laws that consider privacy as a fundamental right of human. Non-technical Perspective, explains privacy laws in western countries in general and then it discusses privacy specifically in eastern countries.

According to Banisar (2018) over 100 countries adopted privacy or data protection laws and around 40 countries have pending bills or initiatives to display the law. There are countries such as the United States, which still have not adopted the law but rather adopted limited sectoral laws in some areas.

A continent specific example of privacy laws is the General Data Protection Regulation (GDPR). GDPR is the European regulation on data protection and privacy which was adopted on 14 April 2016 and came into force on May 25, 2018. GDPR can be helpful since it concentrates on the protection of personal data of an individual not an organization or group of people. Personal data can be any information that identifies persons private information directly or indirectly .

TECHNICAL PERSPECTIVE

In general, any system performs three main operations on information; transfer, store, and process. From the technical aspects of privacy policies, the problems in these operations are the lack of appropriate implementation of privacy principles such as transparency in data transfer, data minimization in data collection and storage, and lawfulness in data processing phases. However, there are general methodologies to make a privacy-friendly system. Although there are no concrete or complete guidelines for engineering a privacy-friendly system, the guidelines have considered some necessary requirements such as legal requirements, users' expectations, and privacy-enhancing technologies.

Well-known research by Ann Cavoukian (2012), Privacy by Design, suggested seven foundational privacy protection principles which had strong impact internationally on reception of privacy's role in systems engineering and operation, as an example, one of the GDPR articles (25th) points it out as "Data Protection by Design and by Default". The principles can be used more as a mindset while engineering a system, but not as a complete guideline for engineering privacy-friendly system.

It is worth to mention that a framework, Privacy Engineering, provides principles and concrete guidelines for engineers and computer scientists to build a privacy-friendly system. The research has described two approaches to engineer privacy; "Privacy-by-Policy" and "Privacy by-Architecture" (Spiekermann & Cranor, 2009). First one focuses on systems that implement notice and choice principles inspired by Fair Information Practices (FIPs) (Rotenberg, 2001). The second one focuses on cryptographic approaches using anonymization and client-side data storage and processing.

Considering all the above-mentioned guidelines, Hoepman (2014), has introduced privacy protection strategies, which covers privacy laws (GDPR and ISO Privacy) and provides accordingly eight privacy design strategies: minimize, hide, separate, aggregate, inform, control, enforce and demonstrate with privacy design patterns and privacy-enhancing technologies.

These strategies mainly have been discussed into two categories with the insight of Privacy Engineering framework. Privacy by Policy have been considered as a "process-oriented" strategy where any process to personal information will be informed, controlled, enforced, and demonstrated to the users and data protection regulations. Privacy by Architecture have been considered as a "data oriented" strategy where any data whether it is going to be stored or transferred will be minimized, hided, separated or aggregated according to the organizational and technical requirements or constraints (Spiekermann & Cranor, 2009).

RECOMMENDER SYSTEMS

With social network development, recommender systems have been increasingly prevalent and have become widely accepted by users. A common risk associated with privacy in recommender systems are methods of personalizing users' preferences according to gathered information, whether explicitly by asking questions or implicitly by tracking their behavioral activities.

A survey of privacy risks associated with personalization-based systems shows that three systems such as social-based, behavioral profiling, and location-based personalization require more attention regarding privacy issues. Obviously, such systems require a lot of data in order to have an efficient and accurate prediction. The problems from technical aspects of privacy policies are the lack of appropriate implementation of privacy principles. Such as transparency in data transfer, data minimization in data collection and storage, and lawfulness in data processing phases (Toch, Wang & Cranor, 2012).

PRIVACY RISK IN RECOMMENDER SYSTEMS

From the privacy perspective, common recommender systems can be considered based on their content personalization methods. The reason is, data practices on users' information which includes their personal data, is done by personalization methods. There are two common personalizing recommenders' categories including explicit personalizing recommenders which use social-based personalization as an example and implicit personalizing recommenders which use behavioral-based personalization as an example to explain the associated risks. Both categories cover three phases of a personalization method which are data collection, user model creation, and adoption, using E.Toch et.al suggested framework (Toch, Wang & Cranor, 2012). The framework explains two or three general methods used by each phase to perform their tasks. As an example, the data collection phase besides gathering the information provided by the user, it also tracks user activities or uses automatic context information.

EXPLICIT PERSONALIZING RECOMMENDERS

Explicit recommender systems can be seen in online commercial services. These services mostly use collaborative filtering methods to recommend their products or services to the customer. The method recommends items according to users' neighbors who have rated similarly other items. By "explicit" it does not mean that the user is completely aware of how gathering and practicing his/her personal data are done. It means that in general, the user is aware that he/she should provide some data including personal data in order to receive a specific service. The willingness of exposing their personal data is not only dependent on the reasons behind users' decisions but also depends on design decisions made by services (Ackerman & Cranor, 1999). Nowadays most commercial-based recommenders are using social-based personalization for a better product or service recommendation. There are different phases of personalization methods in recommender systems such as:

- Data collection: In the social-based personalization method, the main service is to allow users to have a profile in order to communicate with each other while the goal of communication may vary in different social networks differently (Gross & Acquisti, 2005). The collection can be done by the user provided information or by tracking users. Privacy risks are in both of these data collection methods actions. Although the sites provide reasons for asking a demographic related data (for instance, "birthdate" to remind users birthday for his/her friends), which may be useful but not necessary for the main purpose of the service (Bonneau & Preibusch, 2010). Social sites also perform data collection by tracking users' actions. Besides visible actions such as commenting or liking a post or creating content, there are invisible user actions such as browsing a profile page or viewing a photo which can help in collecting more accurate data from users. Facebook priorities "meaningful" conversations between friends and family over stories from publishers, brands, and businesses (Constine, 2018). This feature uses different visible actions of users such as frequency of posts, at the same time it also uses invisible actions of users such as average time spent on content. Tracking users' actions are useful for Facebook users so that they can spend their time more efficiently but for data aggregators, who use social sites as one of their main sources, it can reveal users' personal activities which they kept private(Benevenuto et al., 2009).
- User model creation: Social-based personalization is commonly used to provide services such as social search, personalized recommendation, and targeted advertising. Such services require specific user models to perform recommendations. Therefore social-based systems are using very high-level learning algorithms to learn their users' preferences and create models accordingly.
- Adaption: The adaption phase indicates the distribution of personalized content which can only be to the user, or to user's social network, or to the whole World Wide Web (Toch, Wang & Cranor, 2012). In social-based networks, the adaption is done using all mentioned models.

IMPLICIT PERSONALIZING RECOMMENDERS

Implicit recommender systems can be seen in location-based services. Most mobile applications use location-based services to extract the exact location of users in order to provide more precise recommendation results. Smartphones store sensitive information about the users. Privacy risks do matter because this information is allowed to be accessed by third party applications downloaded from online market stores such as Apple App Store or Google Store. While most data from smartphones are kept secure by the users but location-based information is not easily controllable by the users (Guardian, 2016). Such data are often collected by the applications in the background. There are different phases of personalization methods in recommender systems such as:

- Data collection: Data collection matters when a user who allows an application to access her location information, has no idea whether the application will use her data for purposes that it listed explicitly or it will send her location to other third parties such as location-based service, advertisers, application developer, or to any other entity. (Enck. et al, 2014)
- User model creation: Location-based personalization is not only used by applications that are directly dependent on location information such as touristic recommendations or map applications. There are recommendation systems that use this personalization method as one part of their learning process to have a more accurate prediction. Privacy breach examples related to such models can be scenarios where location-based service providers can observe. Observations include all requests within a specific location from a single user, or all requests during time interval came from a single user within a specific location. Such breaches can help in getting user's sensitive information regarding visited locations, such as a user might have gone to a special hospital or to a place that is related to a special religious party. (Bao. et al, 2015)
- Adaption: different recommender systems are using location-based personalization differently. Thus, the adaption may vary in each system, but from a platform perspective, smartphones are common environments where recommender services can access location-based information via their applications.

DISCUSSION

The important part of any recommender system is its personalization in order to perform predictions. Accuracy in prediction always plays an important role and different personalization methods use different approaches to achieve that goal. While recommender systems could succeed in achieving the desired accuracy, the user's privacy is always violated. It is important to understand that users' satisfaction is not only achieved with accurate predictions. It also requires trust from the system to have constant or permanent user satisfaction. Protecting users' privacy is one way of gaining that trust. Therefore, better user privacy requires user satisfaction from a system, practicing his/her data, in order to use a service.

Different people such as Toch et al. (Toch, Wang & Cranor, 2012), Spiekermann (Spiekermann & Cranor, 2009), and Hoepmann (Hoepman, 2014) define different privacy protection strategies from existing privacy principles using frameworks. According to Toch, Wang, and Cranor (2012) the first framework concentrates on different phases of personalization methods in recommender systems, the second and third frameworks help to support privacy protection throughout the full software development lifecycle (Spiekermann & Cranor, 2009; Hoepman, 2014). Spiekermann's framework is used to cover more general aspects of building a privacy-friendly system. Privacy-by-policy approach for a user-centric design and privacy-by architecture approach for a data-centric approach (Spiekermann & Cranor, 2009).

A detailed guideline for each privacy design strategy is explained by the Hoepmann framework (Hoepman, 2014). It distinguishes three different tools to support the decisions to be made in each phase of the software development phase. Design strategies have been suggested for the first two phases: development and analysis. The design strategies describe the fundamental approach to achieve certain design goals such as privacy protection (Hoepman, 2014). To achieve the privacy protection goal eight strategies have been suggested: minimize, hide, separate, aggregate, inform, control, enforce and demonstrate. For the design phase set of design, patterns are suggested to achieve each design strategy. Typical examples of privacy design patterns are the concept of k-anonymity (Sweeney, 2002), attribute-based credentials (Camenisch & Lysyanskaya, 2001), or mix networks (Chaum, 1981). Finally, for the implementation phase, privacy enhancing technologies are considered to be useful. Privacy Enhancing Technologies (PETs) are used to implement certain privacy design patterns such as u-prove can be used to implement an attribute-based credentials design pattern (Paquin & Zaverucha, 2011).

It is important to know that both Toch et al. (Toch, Wang & Cranor, 2012) and Hoepmann (Hoepman, 2014) frameworks have followed privacy engineering approach based on Spiekermann (Spiekermann & Cranor, 2009) framework.

Besides using some general frameworks and considering privacy protection principles, there are some other general guidelines that concern all sites collecting user's data. In particular, the policies and methodologies that employ with recommender systems should clearly state by individual sites, including the role played by straddles in their datasets and system designs. Some general guidelines to highlight the implications of analyses can be pointed as follow:

- Amount of Data: Some data has more value in differentiating among users. Knowing users' interests in items that have a higher level of diversity among other users' opinions rather than being universally popular, can help in generating much more accurate recommendations. With a VOI (Value of Information): metric, it is possible to calculate the value of given data, and information collection can be optimized with respect to both privacy and recommendation accuracy. Therefore, if the recommender system only keeps a subset of data provided to it, this would require the attacker to know more and increases the cost of an attack. (Shyong, 2006)
- Trust-Based Recommendation: People trust more on recommendations from known people rather than on recommendations generated based on unknown people's similarities. Metrics in trust-based systems use trust propagation and then aggregation; a propagation of transitive computation of trust between users who know each other and then aggregating resulted trust estimations into one final trust value. Trust-based systems use algorithms that are based on implicit trust scores and compute accuracy from past recommendations. As a result, it makes the system less vulnerable to malicious attacks and attack is only possible when the target user has explicitly indicated that he trusts the adversary. (Ricci& Shapira, 2011)
- Privacy-Preserving Cryptographic Protocols (Ciriani et al., 2006): Privacy regulations enforce that identifiable information must be separated or encrypt from other normal information to break associations of sensitive information. However, information encryption has technical limitations, since encryption makes it difficult to have efficient execution of queries and condition evaluation over data. A work by Ciriani et al (2006) has suggested a solution where a trusted application is invoked on request to access sensitive encrypted information from the database which can avoid privacy breaches. The trusted application can use zero-knowledge proof, a protocol that allows a prover to prove a secret without revealing it to the verifier. First, the prover sends a commitment to the verifier, then the verifier asks the prover to open the commitment in a specified way. The commitment can only be opened

correctly when the property of the secret holds (Sasson, 2014).

Nevertheless, complete privacy is not realistic, and that therefore a compromise on minimizing the privacy breaches must be considered. Privacy may also come at the expense of the accuracy of the recommendations. Therefore, it is important to analyze it carefully.

CONCLURION

This paper has argued that users' privacy plays important role in any system practicing their personal data. Although recommender systems enhance user experience in utilizing their services, such enhancements are provided at cost of their privacy loss.

The author began with a general overview of privacy and recommender systems and then introduced privacy from technical and non-technical perspectives, concentrating on personalization phases used in recommender systems. It explained that how such systems violate privacy not only by utilizing users' personal data with explicit mechanisms but also with implicit mechanisms. As an example, social-based recommendations are not only sharing users' private data with user's friends but also they share to the entire World Wide Web. Although various guidelines and frameworks have been published in building a privacy-friendly system such as privacy protection design strategies and privacy engineering approaches but all of them have a general focus on all IT Systems.

The analysis found that user privacy is important in recommender systems. They use the user information, which the user may be aware of or unaware of them. Therefore, users should be careful when using them. Till now there are more general privacy protection guidelines, in the near future, there will be more guidelines for specific IT systems as well.

REFERENCES

A. C. Eli Ben-Sasson,(2014), "Zerocash: Decentralized Anonymous Payments from Bitcoin," May.

A.F. Westin,(1967), Privacy and Freedom. Atheneum.

Banisar, David, National Comprehensive Data Protection/Privacy Laws and Bills 2018 (January 25, 2018). Available at SSRN: https://ssrn.com/abstract=1951416 or http://dx.doi.org/10.2139/ssrn.1951416

Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. (2015). Recommendations in location-based social networks: a survey. GeoInformatica, 19(3), 525-565.

Benevenuto, F., Rodrigues, T., Cha, M., & Almeida, V. (2009, November). Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement (pp. 49-62). ACM.

Bonneau, J., & Preibusch, S. (2010). The privacy jungle: On the market for data protection in social networks. In Economics of information security and privacy (pp. 121-167). Springer, Boston, MA.

Cavoukian, A. (2012). Privacy by design: origins, meaning, and prospects for assuring privacy and trust in the information era. In Privacy protection measures and technologies in business organizations: aspects and standards (pp. 170-208). IGI Global.

Camenisch, J., & Lysyanskaya, A. (2001, May). An efficient system for non- transferable anonymous credentials with optional anonymity revocation. In International Conference on the Theory and Applications of Cryptographic Techniques (pp. 93-118). Springer, Berlin, Heidelberg.

Chaum, D. L. (1981). Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 24(2), 84-90.

D. F. Shyong K. (2006). "Tony" Lam and John Riedl, Do You Trust Your Recommendations? An Exploration of Security and Privacy Issues in Recommender Systems, vol. 3995. Springer, Berlin, Heidelberg.

D.J. Solove,(2005) , "A Taxonomy of Privacy," Univ. of Pennsylvania Law Rev., vol. 154.

Enck, W., Gilbert, P., Han, S., Tendulkar, V., Chun, B. G., Cox, L. P., ... & Sheth, A. N. (2014). TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones. ACM Transactions on Computer Systems (TOCS), 32(2), 5.

The Guardian, "Is your private phone number on Facebook? Probably. And so are your friends'," The Guardian, 2016. [Online].

Gross, R., & Acquisti, A. (2005, November). Information revelation and privacy in online social networks. In Proceedings of the 2005 ACM workshop on Privacy in the electronic society (pp. 71-80). ACM.

L. Sweeney. (2002)“k-anonymity: a model for protecting privacy.” International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 557-570.

L. R. Francesco Ricci and P. B. K. Bracha Shapira, (2011), Recommender Systems Handbook. New York Dordrecht Heidelberg London: Springer.

I. Altman,(1975), The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding. Brooks/Cole.

Josh Constine,(2018) “Facebook feed change sacrifices time spent and news outlets for ‘well-being,’” Tech Crunch, 12-Jan-2018. [Online]. Available: https://techcrunch.com/ 2018/01/11/facebook-time-well-spent/.

M.S. Ackerman, L.F. Cranor, and J. Reagle,(1999) “Privacy in ECommerce: Examining User Scenarios and Privacy Preferences,” Proc. First ACM Conf. Electronic Commerce, pp. 1-8, http:// doi.acm.org/10.1145/336992.336995, Nov.

Rotenberg, M. (2001). Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get). Stan. Tech. L. Rev., 1.

Paquin, C., & Zaverucha, G. (2011). U-prove cryptographic specification v1. 1. Technical Report, Microsoft Corporation.

Hoepman, J. H. (2014, June). Privacy design strategies. In IFIP International Information Security Conference (pp. 446-459). Springer, Berlin, Heidelberg.

Toch, E., Wang, Y., & Cranor, L. F. (2012). Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems. User Modeling and User-Adapted Interaction, 22(1-2), 203-220.

Whitman, James Q., (2004), "The Two Western Cultures of Privacy: Dignity versus Liberty". Faculty Scholarship Series. Paper 649.http://digitalcommons.law. yale.edu /fss papers/649

Spiekermann, S., & Cranor, L. F. (2009). Engineering privacy. IEEE Transactions on software engineering, 35(1), 67-82.

S. D. C. di V. Valentina Ciriani, S. J. Sara Foresti, and P. S. Stefano Paraboschi,(2006), “Fragmentation and Encryption to Enforce Privacy in Data Storage,” Italian Ministry of Researh