# Privacy-Preserving Data Mining in Peer to Peer Networks

Ibrar Hussain
School of Science & Computing,
University of Salford,
Manchester,UK
i.hussain@pgt.salford.ac.uk

Maria Irakleous
School of Science & Computing,
University of Salford,
Manchester,UK
m.irakleous@pgt.salford.ac.uk

Mubashir Ashraf Siddiqi
School of Science & Computing,
University of Salford,
Manchester,UK
m.a.siddiqi@pgt.salford.ac.uk

Mohamad Saraee
School of Science & Computing,
University of Salford,
Manchester.UK
m.saraee@salford.ac.uk

## ABSTRACT

*In recent years, privacy-preserving data mining has been studied extensively, due to the wide increase of sensitive information on the internet. A number of algorithms and procedures have been designed, some of which are yet to be implemented, but a few of them are actually employed in the form of software systems to preserve the privacy of users, and the content in peer-to-peer networks. Privacy issues are becoming widely recognized when using peer-to-peer networks. In this paper, we provide a review of the privacy-preserving data mining techniques used in order to overcome privacy issues.*

*We discuss methods of sanitization, data distortion, data hiding, cryptography and the data mining algorithm KDEC. Further discussion involves data transfer using proxy techniques, creating social communities among peer-to-peer users forming trusted peers. These techniques have shown to administer the issue of preserving data however show lack of scalability and performance. We design a framework to perform a comparison study on the techniques shown above and present the results with some recommendations of how we think the issues could be unraveled.*

## Keywords

Peer-to-Peer (P2P), Data mining, Privacy Preserving Data Mining (PPDM), Distributed Data Mining (DDM)

## 1. INTRODUCTION

File sharing and P2P is seen as forbidden by many people due to the simple fact that the holders of copyright are not acknowledged in the form of payment for their materials. There is a different kind of network that is becoming an increasingly common experience in the average internet user's daily life, which is called P2P networks. It simply connects users directly instead of through a central point. The combined bandwidth is used to transfer data [1]. Furthermore it is seen that viruses are available on most of the programs in the form of music or video files which in turn can damage the user's computer [2]. It is clear that there is no need for a client and server as each node acts as a peer on the network. Such networks are gaining popularity in applications such as file sharing, e-commerce, and social networking, many of which deal with distributed data sources that can benefit from data mining [3]. One of the main issues with a P2P relationship is of existing privacy-preserving data mining (PPDM) techniques.

Data mining is a recently emerging technique, where large volumes of data are gathered. Statistical methods show that data mining techniques use concepts and algorithms such as association rule learning or inductive-rule learning in the form of decision trees.

While data mining is a technique having many advantages, it holds the most common problem of privacy [4]. The increase in the transfer of personal information has often led to identity theft however it has been suggested that data mining will indulge a huge debate in the coming years. It has been argued that people should be given the right to choose whether their details are stored in a database. Many people do not know how data mining works and how it can help in the long run.

P2P networks are in essence, hugely related to [5] distributed data mining (DDM), which is where data is distributed into several databases, making a centralized processing of this data, thus protecting it from security attacks. This also deals with the problem of data analysis in environments with distributed data, computing nodes, and users.

In summary it is also shown that users who use peer to peer also leave themselves open to unauthorized access, causing an invasion of privacy. This can happen when your computer is left with an active internet connection for a long period of time. An example of an occurrence could be if a user wants to download a large file taking up to 3 hours to complete. The user would start the download and return when the file was downloaded. This leaves the user open to attacks and could result in catastrophic events. Therefore there is a necessity of looking at ways in which data mining helps and the techniques are used today to protect users.

It is without doubt that we need to retrieve knowledge from peer to peer applications, which implies the fact that data mining is needed. However as privacy is a huge concern, we need to use PPDM.

As there has not been an exhaustive attempt to acknowledge research in relation to PPDM we will do a survey of existing work looking at the issues it brings. At the end we

will have a discussion to compare and contrast the different techniques found.

The rest of the paper is organized as follows. In Section 2 we will thoroughly examine the techniques that have been used in P2P networks by carrying out a survey on a series of papers. Section 3 discusses the concepts of the various methods analyzed in this paper. We converse the ideas put forward exhausting the advantages and disadvantages, followed by a short section suggesting our views on how the issue of privacy could be overcome. Section 4 concludes the paper with several directions for future work.

## 2. Overview of Recent Work

This section presents a view of the recent work that has been researched in relation to privacy preserving and P2P networks.

### 2.1 Peer-to-Peer Data Mining, Privacy Issues, and Games

Research conducted by Das et al (2007) elaborated on the fact that PPDM can be divided into two groups, data hiding and rule hiding. They gave an overview of DDM and algorithms for P2P environments along with explaining the privacy concerns of existing privacy preserving multi-party data mining techniques.

Their work presented a motivational application emphasizing the detail to which preserving the privacy of users was important. This application used the frequency of the web domains a user visited during a specific period of time as the users profile vector. It was increasingly obvious that this method had its weakness where it involved users having to share their actual browsing data.

Taking this into account Liu et al (2006) used cryptographic secure inner product protocols to compute the inner product between two users profile vectors. This was believed to have helped preserve user's private data; however there was still room for improvement. To explain this, there is no control over the user's behavior and are also not monitored. The points put forward by Liu et al (2006) are seen to be computationally very intensive [7], expensive and not scalable in any way.

Camara et al (2009) supported this judgment in explaining how excessive overhead communication held a key role in PPDM. They proposed a new scalar product which aimed to reduce this communication. Therefore it was empirical that a more sheltered approach was taken into ensuring privacy in a P2P setup.

Moving further the study conducted by Das et al (2007) showed innovation by formulating a game theoretic approach to PPDM which did not suffer from the problems described above. This method presented PPDM algorithms designed as games. This is where they modeled the large-scalemulti-party mining applications as games. This is where each participant tries to maximize its benefit by choosing the strategies during the PPDM process. They looked at multi-party PPDM in a more realistic scenario not looking at them to be well behaved. From looking at this they implemented a solution and performed multi-agent simulations in order to study the behavior of the agents.

The deployment could however be criticized as there is a need for a highly scalable and efficient algorithm for data integration. It is argued that existing PPDM algorithms assume that the parties are well-behaved and they abide by the protocols as expected. This piece of research tried to offer a more realistic formulation of this issue by maximizing its own objective. However from looking at this study overall it only offers a new approach using a different approach. Scalability is the main stumbling block for the cryptographic PPDM. It also does not address the question of whether the disclosure of the final data mining result may breach the privacy of individual records [9].

In summary although there has recently been studies on Distributed Data Mining (DDM) as a possible solution, proposed DDM algorithms cover a small portion of the problem space and lack a theoretical proof of convergence. A possible solution to this may be to offer a layered data-gathering and computing infrastructure.

### 2.2 Inference Attacks in Peer-to-Peer Homogeneous Distributed Data Mining

Further research conducted by da Silva et al (2006) analyzed the potential threats to data privacy in a P2P agent-based distributed data mining scenario. They also discuss interference attacks which could compromise data privacy in a P2P distributed clustering scheme known as KDEC. It is seen that in the last decade the extraction of patterns or huge centralized datasets and DM has become admired. However as the internet holds an extensible factor in encouraging the issue of privacy, through the means of data sharing, many research projects have been undertaken.

This paper looks at previous attempts addressing the privacy issue. This includes the *sanitization* and the *data distortion* approaches. However, preventing interference attacks in open environments is suggested to be difficult if not impossible. As already explained in the first paper this is to do with the reason of scalability and within this paper concludes due to the fact of untrustworthiness of involved parties. Secure Multi-party Computation (SMC) is also another method which involves sharing minimum information between involved parties.

It has been shown in studies that SMC can be applied to the data mining process with a few changes of the original idea with respect to data input size [11], [12]. These approaches have not been seen to solve the problem of privacy. This is

where the introduction of KDEC appears, an algorithm which is a distributed clustering scheme. This can be taken as a solution to homogenous distributed data clustering (DDC). This is when the clustering specification is based on a nonparametric kernel density estimate of the data.

The kernel density estimates:

- Additive for homogeneous distributed datasets,
- Can be transmitted in sampled form in order to hide the data points, which are otherwise explicit in the representation of a kernel estimate.

In summary this approach takes advantage of multi-dimensional information sampling to minimize communications among sites this increasing privacy. However the accuracy of the algorithm could be improved to depict further possible weaknesses of the KDEC scheme. This would also act as a way of providing countermeasures to such attacks. Again this paper can be subject to criticism as the algorithm used is not highly scalable, nevertheless it does have a glimpse of efficiency.

The battle against users that share copyright material over the Internet using networks like Gnutella is intensively going, many in the form of cyber attacks. This increases the number of concurrent connections and anyone with access to the network has automatic authorization for sharing folders.

So far both the papers have been able to suggest methods of preserving the privacy of data, however as they have weaknesses there is a necessity to look into a more rigorous approach which is more appealing and hold less risk of becoming just 'another failure'. We further discuss research which has endeavored to do this and point out the key areas in overcoming the issue of privacy in P2P systems.

## 2.3 Towards Data Mining in Large and Fully Distributed Peer-to-Peer Overlay Networks

Research conducted by Kowalczyk et al (2003) targeted the problem of analyzing data that is scattered over a huge and dynamic set of nodes, where each node is storing possibly very little data but where the total amount of data is immense due to the large number of nodes.

Their work presented distributed algorithms for the effective calculation of basic statistics of data using the newscast model of computation [14] and also demonstrated how to implement basic data mining algorithms based on these techniques. The suggested techniques described were efficient, robust, and scalable and they preserved the privacy of data.

According to Kowalczyk et al (2003), because the Internet is already being used to support huge, fully distributed peer-to-peer overlay networks that contain millions of nodes for the purpose of file sharing and information dissemination, the motivations behing Distributed Data Mining (DDM) include the optimal usage of available computational resources, privacy and dependability by eliminating critical points of service.

The constraints adopted on the distribution of data and the elements of the network were that all nodes are allowed to hold as few as one single data instance and there is no limit on the number of nodes in the network. Each pair of nodes can communicate directly which holds if the nodes are on the Internet with an IP address. Kowalczyk et al (2003) concentrated on data privacy and the dynamic nature of the underlying network where nodes can leave the overlay network and new nodes can join it.

Using the newscast model of computation [13], there exists the advantage that the applications of the newscast model of computation inherit the robustness and scalability of the model and can target all the kinds of distributed networks. The peers never communicate directly to each other but through a news agency, that although the news agency plays the role of a server in this model, it is a purely virtual entity and the actual implementation of its functionality at the protocol level is a fully distributed peer-to-peer solution.

Taking this into consideration, the peers are receiving only the contents of news items but no other information about the sender, therefore the system stays completely anonym and privacy of the peers is not violated. The achievement is that the origin of a given item is hard to track down, as the protocol that implements this model can effectively act as a 'remailer'.

## 2.4 Client-side Web Mining for community Formation in Peer-to-Peer Environments

Research conducted by Liu et al (2006) presented a framework for forming interests-based peer-to-peer communities using client-side web browsing history with the use of order statistics-based approach to build communities with hierarchical structure.

They have taking into consideration the privacy concerns of the peers and have adopted cryptographic protocols to measure the similarity between them without disclosing their personal profiles.

Their work addressed the problem of forming interest-based communities in a Peer-to-Peer environment using the attribute similarity-based approach proposed by Khambatti et al (2002) where each peer has a set of attributes called profile vector. Liu et al (2006) used this approach but also extended it by giving a weight to each interest in the profile vector to show its importance. Instead of simply checking the intersection of attributes, they quantitatively computed

the similarity between profile vectors using scalar product and order statistics-based algorithm that can tell how similar a pair of peers are to each other in the whole network.

In their framework they provided the peer with a two-level privacy protection:

- The first level allows the peer to explicitly filter out extremely private sensitive interests by assigning zero weights to the corresponding concepts in the profile vector and,
- The second level protection relies on the notion of cryptographic secure multi-party computation (SMC) [14].

The advantage of using SMC-based protocol is the guarantee that neither party would know each other's actual input, namely, the actual profile vector.

Taking into consideration the user anonymity that aims to offer the users privacy protection by letting them hide their identities from the communicating peers or from malicious eavesdroppers and the protection of data privacy that aims to hide the sensitive information owned by a peer from being disclosed they have used the Paillier cryptosystem for public-key cryptography and computed the scalar inner product between two users profile vectors. Using this approach they have preserved the user's private data without compromising the privacy of their own profiles.

### 2.5 Trust-Based Privacy Preservation for Peer-to-Peer Data Sharing

Research conducted by Yi Lu et al (2009) on PPDM came up with another idea of how to preserve the privacy of the user's identity in P2P environment. They proposed the technique of Proxy to be used in P2P file sharing environment and through this method the identity of the user can be protected and the privacy of the data can be maintained.

Their research extends the earlier work conducted by Bhargave et al (2002), and Lilen et al (2003), which addressed the issues of hiding the identity of the requester and creating trust based community where each peer creates relationships with another peer also known as "Buddies". This acts as a proxy for the requester, so that the supplier does not know the identity of the requester.

The issues concerning privacy of the requester is maintained by involving a trusted buddy between the requester and the supplier. The trusted buddy forwards the request to a number of suppliers and takes back the response to the requester. The supplier does not know the identity of the requester; hence the privacy of the requester is achieved.

Once the privacy of the requester is achieved through the buddy acting as a proxy raises the issue of trust worthiness of the buddy and protection of the data handle and protection of the content.

The issue of protecting the data handle can be achieved by introducing the hash value which requester calculates before sending the request, hash value is revealed partially to the supplier if the partial hash code matches, then the supplier can create a reply with a public key of supplier and replies it to the buddy the buddy then forwards it to the requester where requester can encrypt the complete request and the content with the public key and send it to the buddy. Buddy then has to sign the packet so that it ensures that the packet is not generated by the buddy itself but by the supplier and then it can be forwarded to the supplier. This approach solves the problem of protection of data handle and the content.

The main issue now is of the trust worthiness of the buddy and to calculate it in the dynamic environment. A model is set to calculate the trust worthiness of the buddy based on its behavior and other peers' recommendations. The peer's behaviors, such as keeping a secret while being a proxy forwarding requests in a timely fashion, buffering data to improve streaming-capacity etc will affect the trust worthiness metric. Communication principles, such as Kalman filtering [27] are applied to build a trust model as a multivariate, time-varying state vector that utilizes past information to predict future performance.

The problem with this model is that the whole system is dependent on the peer's relation to the buddy it is quite possible that even with a large peer population, the overall capacity is low because of the lack of buddy relations in the system. It is also possible that a small network of buddies of many peers might become over loaded because they are involved in too many data sharing sessions.

More experiments are still being conducted in prototype environment to check the limitation of the system in a large scale network.

### 2.6 Privacy-Preserving P2P Datasharing with OneSwarm

Recent work conducted by Isdal et al (2009) describes the design, implementation, and experience with OneSwarm, a new P2P data sharing system that provides users with explicit, configurable control over their data: data can be shared publicly or anonymously, with friends, with some friends but not others, or only among personal devices. The main concern is to reduce the performance cost of privacy while maintaining the scalability and performance of the system.

Oneswarm provides features, which enables users to act as a replica for sharing without attribution using an overlay consisting of OneSwarm peers only. This Over lay act as a

mix using source-address rewriting and multi-hop overlay forwarding to unclear the identities of a path's source and destination [27]. OneSworm also provides users the ability to download the data using only anonymizing paths to preserve her privacy from third-party monitoring, at the same time they can advertise their files explicitly to friends.

OneSworm provides 3 different types of privacy for file sharing 1. Public distribution where all the files are shared publically and everyone on the system can access the files. 2. With Permission: this type of privacy allows users to share files among restricted/ allowed users only 3. Without Attribution: data shared without attribution is located using privacy-preserving keyword search, and data transfers are relayed through an unknown number of intermediaries to obscure source and destination. This type of distribution is appropriate for sensitive material.

OneSwarm provides capability for the users to add other peers as friends by sharing a cryptographic key identity which identifies users and their friends. The key can be shared in different ways. 1. Manually over local area network. 2. Users can email the invitations to friends and in one time process the keys are shared. This information is stored in a special Distributed Hash Table (DHT).

OneSwarm has high concerns about performance. Therefore the search method of OneSwarm does not rely on shortest path search like other P2P systems. OneSwarm enhances its search to distant network in order to avoid congestion and over load on the nearest peers. And it tries to find multiple paths in case one of the path is lost to it can carry on sharing through other provided paths. Rather than connecting directly to peers, OneSwarm connects through overlay paths. Each overlay path is treated as a virtual peer, even those that terminate at the same endpoint. It follows the keep-alive protocol which checks the status of each path after 30seconds and its dead then it changes to the alternative path.

OneSwarm maintains the security over the network attacks by following different protocols such as Skewed object popularity motivates popularity-aware search, Long paths motivates multipath downloads from a single source, a resilient core improves availability but requires adaptation to congestion, Bootstrapping is crucial since many users have few trusted links.

To conclude One Swarm is the first ever Peer-to-Peer system that provides scalability and Performance while maintaining privacy of the users as well.

## 3. DISCUSSION

Research has shown that there has not been a methodical attempt to acknowledge research in relation to PPDM, therefore there is a need to examine the techniques that are used to preserve the privacy in P2P systems. We have therefore done a survey of existing work taking into consideration the basic concepts that are used.

There has been a rise in the use of cryptography to facilitate the way data is communicated between peers, meaning that data is mathematically manipulated for the purpose of its security, so that information is hidden from anyone for whom it is not intended, even those who can actually see the manipulated data [29]. Research shows remarkable results have been achieved while using cryptography, thus enhancing efficiency demonstrating their relevance to privacy preserving computation of data mining [24]. The points put forward in (2.4), show how cryptography was used, and the way in which this technique was implemented to hide data. This method is seen to increase security, and the private keys do not need to be transmitted ensuring the identity of a peer is correct. However this procedure is rather slower and lacks scalability therefore affecting the performance of the system which is seen as one of the biggest problems in data mining.

Another method seen is sanitization and data distortion which work hand in hand where sensitive information is kept safe. The risk of unexpected information leaks is increasing, so it is imperative that a technique is used to prevent this. In section (2.2) we discussed how the use of sanitization was used and although this technique is efficient, it requires a significant amount of data distortion to preserve privacy, therefore it was not helpful in preventing interference attacks. This moves on to suggest that the KDEC algorithm was used to overcome this. This approach was also seen to be a weaker method of preserving the privacy of data. This elaborates on the fact that scalability and efficiency played a huge factor.

| Techniques Used | Paper 1 | Paper 2 | Paper 3 | Paper 4 | Paper 5 | Paper 6 |
|---|---|---|---|---|---|---|
| Sanitization | Yes | ----- | ----- | ----- | ----- | ----- |
| Data Distortion | Yes | ----- | ----- | ----- | ----- | ----- |
| Secure Multi-party Computation | Yes | ----- | ----- | ----- | ----- | ----- |
| KDEC algorithm | Yes | ----- | ----- | ----- | ----- | ----- |
| Distributed algorithms | Yes | ----- | Yes | ----- | ----- | ----- |
| Cryptography | ----- | Yes | ----- | Yes | Yes | ----- |
| Buddies Technique | ----- | ----- | ----- | ----- | Yes | ----- |
| OneSwarm | ----- | ----- | ----- | ----- | ----- | Yes |

The newscast model of computation developed during the DREAM project [13] for distributed evolutionary computing frameworks provides the advantage of effective and reliable multicasting, large-scale distributed file sharing and resource discovery and allocation. When compared to systems that provide key-based routing and searching [30], [31], the newscast model solves the problem of content-

based searching due to is disseminative nature, but at the cost of higher resource usage. This model is characterized by robustness, fault tolerance, decentralized approach involving peer-to-peer networking with fast and not overloaded network communication transfer lines thus enhancing and preserving the privacy of data shared among peers. In section (2.3) we have discussed the use of the newscast model of computation and it is seen that this approach is scalable and robust, allowing the peers in the overlay network to remain completely anonym thus ensuring that privacy cannot be violated.

Secure multi-party computation (SMC) techniques [32] have recently emerged as one of the answers to privacy preserving distributed data mining. As discussed in section (2.4) cryptographic SMC is a reliable method that provides privacy protection by evaluating a function of the private inputs from two or more parties such that no party can learn anything beyond what can be implied from the party's own input [14]. Even if the SMC based methods provide efficiency and privacy protection, the use of those methods do not scale efficiently for large amounts of data.

Scalability and efficiency are the major issues of the peer-to-peer system. Trust-based privacy preservation explained in section (2.5) used the method of proxy and tust based relationship between users to hide the identity of the users from outside the network and also within the network. This system proposes that the identity of the requested and of the supplier should not be known to each other hence there should be a middle man called "buddy" that can perform the request handling on behalf of requester and sender this will achieve privacy of the users. Issues related to hiding the data through cryptography where public keys can be shared between the sender and the requester to make sure that the data packets are not generated by the buddy, but by the user should be embedded within the same system so as to achieve a higher level of security. The problem associated with the systems include the trust-worthiness of the buddies and this has to be computed at the real time environment so this makes it quiet challenging and it effect the scalability and performance of the system, different algorithms and model presented in [33.34] are presented to calculate the trust worthiness of the buddies. Prototypes for Trust-Enhances Role Assignment [35] and other supporting software products are being developed which will then determine course of action to achieve the efficiency and accuracy of the proposed system. Section (2.5) still had some defect left out which are further discussed in section (2.6) by presenting a software solution OneSwarm.

OneSwarm has so far accomplished the task by implementing a new peer-to-peer system that wraps up the privacy of users by providing the user a very easy and enhanced level of privacy for its content, each user can define different level of privacy of each content shared. OneSwarm also provides the feature to the users for managing their friends by sharing the public key so that the files could be shared and accessed reliably and efficiently. The system has effectively covered the aspect of performance of the system by sharing file list message when they connect to each other. File list messages are compressed XML which contain the information about the name size description shared date shared which makes it easier for user to understand. The path that the OneSwarm follows for the search is not the nearest system like other peer-to-peer systems. To avoid traffic on the system it forwards the search to the user which is sitting idle so each user also maintains the idle time to avoid duplication of forwarded search. The system so far is one of its kind peer-to-peer systems which has achieved the privacy of the users as well as maintained the efficiency, performance and scalability.

Refer to table 3.1, for a summary of the privacy preserving methods that each technique analyzed in this paper utilizes.

## 4. CONCLUSIONS

We have provided an overview of the most essential algorithms and procedures used for privacy-preserving data mining to overcome the privacy issues in peer-to-peer networks. We have focused our comparison on the methods of sanitization, data distortion, data hiding, cryptography and the data mining algorithm KDEC, while presenting the major advantages and disadvantages of each. The techniques involved the comparison of theorists summarizing the procedures that have been put in place to.

We are hoping to see in the future, techniques that will resolve once and for all the problem of privacy in P2P networks while maintaining their performance and scalability with the ultimate purpose of sharing data securely.

## 5. REFERENCES


[1] Frank J Klein. (2009). *The Good and Bad of P2P Networking.* Available: http://www.relativitycorp.com/networkdesign/article8.html. Last accessed 16 March 2010.

[2] Bhaduri, K., Kamalika, D., and Kargupta, H. 2007 Peer-to-Peer Data Mining, Privacy Issues, and Games. Universiry of Maryland, Baltimore County.

[3] Wikidot. (2009). *Disadvantages* . Available: http://peer2peer.wikidot.com/disadvantages. Last accessed 16 March 2010.

[4] Exforsys Inc. (2009). *Data Mining - Data Mining Privacy Concerns.* Available: http://www.exforsys.com/tutorials/data-mining/data-mining-privacy-concerns.html. Last accessed 16 March 2010.

[5] ai.cs.uni-dortmund.de. (2009). *Distributed computing.* Available http://www-ai.cs.uni-dortmund.de/auto?self=$ejr31cyc. Last accessed 18 March 2010.

[6] K. Liu, K. Bhaduri, K. Das, P. Nguyen, and H. Kargupta. Client-side web mining for community formation in peer-to-peer environments. *SIGKDD Explorations*, 8(2):11–20, 2006.

[7] B. Awerbuch, A. Bar-Noy, N. Linial, and D. Peleg. Compact distributed data structures for adaptive network routing. *Proceedings of the 21st ACM Symposium on the Theory of Computing (STOC)*, 1989.

[8] F. Camara, S. Ndiaye, and Y. Slimani. *A Secure Protocol to Maintain Data Privacy in Data Mining*. 2009.

[9] E. Alexandre, and G. Tyrone. *Privacy-Preserving Data Mining*. 2009.

[10] J. C. da Silva, K. Matthias, L. Stefano, and M. Gianluca. *Inference Attacks in Peer-to-Peer Homogenous Distributed Data Mining*. 2006.

[11] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M.Y. Zhu, `Tools for privacy preserving data mining', *ACM SIGKDD Explorations*, **4**(2), 28.34, (2002).

[12] Benny Pinkas, `Cryptographic techniques for privacy-preserving data mining', *ACM SIGKDD Explorations*, **4**(2), 12.19, (2002).

[13] M. Jelasity and M. van Steen, *'Large-scale newscast computing on the Internet*. Technical Report IR-503, Vrije Universiteit Amsterdam, Department of Computer Science, Amsterdam, The Netherlands, Oct 2002. http://www.cs.vu.nl/globe/techreps.html

[14] M. Khambatti, K.D. Ryu and P. Dasgupta. *'Effiicient discovery of implicitly formed peer-to-peer communities'*. International Journal of Paraller and Distributed Systems and Networks, 5(4):155-164, 2002.

[15] T. Chalfant, "Role based access control and secure shella closer look at two solaris operating environment security features,"Sun Microsystems Blueprint, June, 2003.

[16] E. Lupu and M. Sloman, "Reconciling role based management and role based access control," in *Proceedings of Second ACM Workshop on Role Based Access Control*, 1997, pp. 135–142.

[17] A. Singh and L. Liu, "Trustme: Anonymuous management of trust relationships in decentralized P2P systems," in *Proceedings of The Third IEEE International Conference on Peer-to-Peer Computing*, 2003.

[18] K. Aberer and Z. Despotovic, "Managing trust in a peer-2-peer information system," in *CIKM*, 2001

[19] B. Bhargava and Y. Zhong, "Authorization based on evidenceand trust," in *Proc. of International Conference on Data Warehousing and Knowledge Discovery (DaWaK'02)*, Aix-en-Provence, France, September 2002.

[20] B. Bhargava, "Vulnerabilities and fraud in computing systems,"in *Proc. of International Conference on Advances in Internet, Processing, Systems, and Interdisciplinary Research (IPSI'03)*, Sveti Stefan, Serbia and Montenegro, October2003.

[21] L. Lilien and A. Bhargava, "From vulnerabilities to trust: A road to trusted computing," in *Proc. of International Conference on Advances in Internet, Processing, Systems, and Interdisciplinary Research (IPSI'03)*, Sveti Stefan, Serbia and Montenegro, October 2003.

[22] L. Lilien, "Developing pervasive trust paradigm for authentication and authorization," in *Proc. of Third Cracow Grid Workshop (CGW'03)*, Krakow (Cracow), Poland, October 2003.

[23] D. L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM, 24(2):84–90, 1981.

[24] B. Pinkas. *Cryptographic techniques for privacy-preserving data mining.* ACM SIGKDD Explorations Newsletter, v. 4, pp. 12-19. 2002.

[25] L. Lilien and A. Bhargava, "From vulnerabilities to trust: A road to trusted computing," in *Proc. of International Conference on Advances in Internet, Processing, Systems, and Interdisciplinary Research (IPSI'03)*, Sveti Stefan, Serbia and Montenegro, October 2003.

[26] L. Lilien, "Developing pervasive trust paradigm for authentication and authorization," in *Proc. of Third Cracow Grid Workshop (CGW'03)*, Krakow (Cracow), Poland, October 2003.

[27] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 8, pp. 35–45, 1960

[28] D. L. Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM, 24(2):84–90, 1981.

[29] MXC Software (2007). What is Cryptography? Available: http://mxcsoft.com/Cryp_What%20Is%20Cryptography.htm Last accessed 27 March 2010

*[30]* Steven D. Gribble, Alon Halevy, Zachary Ives, Maya Rodrig, and Dan Suciu, "What can databases do for peer-to-peer?," in *Proceedings of* the 4th International Workshop on the Web and Databases (WebDB'2001), May 2001.

[31] Matthew Harren, Joseph M. Hellerstein, Ryan Huebsch, Boon T. Loo, Scott Shenker, and Ion Stoica, "Complex queries in DHT-based peer-to-peer networks," in Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02), Mar. 2002.

[32] Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game - a completeness theorem for protocols with honest majority. In *19th ACM Symposium on the Theory of Computing*, pages 218.229, 1987.

[33] Y. Zhong, "Formalization of dynamic trust and uncertain evidence for user authorization," PhD Thesis, Department of Computer Sciences, Purdue University, 2005.

[34] N. Li, W. H. Winsborough, and J. C. Mitchell, "Distributed credential chain discovery in trust management," *Journal of Computer Security*, vol. 11, no. 1, pp. 35–86, February 2003.

[35] "Trust-enhanced role assignment (TERA) prototype," *http://raidlab.cs.purdue.edu/zhong/NSFTrust/*.