OneSwarm builds trusted links through social network peers, instead of relying only on a directory service such as a "Tracker" that gives information to the peers about the file. OneSwarm users are free to control the tradeoff between performance and privacy by managing the level of trust.

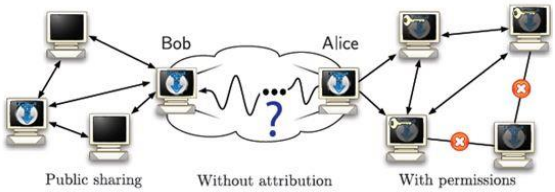

Figure (4). Cases for data sharing by OneSwarm. Source: [9]

There are three cases for OneSwarm described by [9] and shown in Figure 4, the first one is public distributed data in this case the data is not private, and direct transfers between a large set of replicas yield. The second is sharing data with permissions limits access. The last one, data shared without attribution is accessible by everyone. In public distribution anyone in the network can download file free, all data is not private, and serves as fully backwards compatible BitTorrent client. With permission case only users with permission can download files, uses persistent identities to define per file permission, this case allows all acceptable users to recognize one another. While without attribution case is depend on obscuring attribution of source and destination, it uses privacy preserving keyword search, data is relayed through unknown number of intermediaries, and it is for sensitive material.

The topology for OneSwarm the users define the links by exchanging public keys, this identifies each user and creates direct encrypted P2P connections, also OneSwarm uses social graph and community server for key distribution, Distributed Hash Table (DHT) serves as name resolution service, each client maintains encrypted entities advertising their IP address and port to authorized users, the topology is used for each transfer. In each transfer each OneSwarm client restricts direct communication to a small number of persistent contacts and locates different data sources using object lookup through overlay, this topology is used to enhance privacy, while to enhance performance in OneSwarm protocol, multiple paths to each data source are used.

Linking peers with trust relationships is explained by Isdal et al. [9] it uses 1024 bit RSA (Rivest-Shamir-Adleman cryptosystem) public / private key pair which is generated in installation phase, public key serves as its identity among friends, manual key sharing between two users; the automatic key sharing discovers and exchange keys over local area network or by email invitation to friends. Managing untrusted peers by private community server and public community server, the private is to maintain a list of registered users and to provide authorized subscribes with a current set of public keys, the public is to allow new users to easily obtain a set of untrusted peers. Identity in OneSwarm protocol are managed by the DHT which contain of hashed IP and port, entries for a client encrypted with the public key, each entry is indexed by 20 byte randomly generated.

Naming and locating data in OneSwarm used Secure Sockets (SSLv3) for connection as [9] say, file list messages is exchanged on first connection then compressed XML attributes which contain name, size and other meta data for particular peer. Shared files are named using 160 bit SHA-1 hash, for public data user obtains hashes from email, websites and keywords search, while for private data user must obtain both hash and key used for decryption of data. The risk in OneSwarm model as Isdal et al. [9] describe, the attacker can join with limited number of nodes, also can check the traffic flow to/from, also may sniffing, modify or injected data. Limiting hacker to snoop in from by not assigning peer dynamically, also defining trusted and untrusted links to keep the information private, end to end path between users change rapidly helps to prevent hacking using historical data. Isdal et al. [9] add that preventing timing attack by search queries and responses are forwarded after adding a random delay to inhibit calculation of round trip time (RTT) to infer proximity, preventing correlation attack by having limited view of the overlay and cannot control path setup beyond directly connected neighbors, attackers could use this to correlate performance with ongoing transfers, finally preventing collusion attack by search queries and responses are forwarded probabilistically, making it very hard for directly connected colluding peers to infer source of data or monitor habits.

- **OneSwarm Data Sharing Model Strengths**

OneSwarm provides flexibility for the user to manage the level of privacy for file sharing, incorporation of social network for building P2P file sharing network, and reduce cost of privacy.

- **OneSwarm Data Sharing Model Weaknesses**

There are Delayed responses to queries from untrusted peers.

Based on the comparison between the three models, the Capability-based Access Control model has disadvantages, mainly: there is no fixed method for translation, difficulties in integrity control, cannot prevent the user from keeping and distributing the shared data, and decentralized control. These disadvantages make the implementation of the model hard, concerning semantic privacy preserving model overcomes the previous disadvantages, and provides

data integration, secure sharing through authorized view, in addition, each organization enable data sharing without affecting its clients, while OneSwarm data sharing model provides flexibility for the user to manage the level of privacy for file sharing, and reduces cost of privacy but it has delay in response.

4. Privacy and Encryption Techniques

Different encryption techniques used to get secure communictions such as (RSA) Rivest-Shamir-Adleman cryptosystem [18], Rabin [22], Huffman coding [23]. Understanding encryption technique will help to develop more technique to protect information. This section provides a literature review for these techniques.

4.1. RSA Encryption

Encryption is the most used techniques among transactions over insecure channels that used to hide data or information by transforming it into a code, and to protect data being transferred between devices. RSA (Rivest- Shamir- Adleman) is the most widely encryption technique that used, Ron Rivest, Adi Shamir, and Leonard Adleman developed this technique in 1977 which relies on multiplication and exponentiation. RSA is built from two large prime numbers; these prime numbers are manipulated to give a public key and private key. Anyone can encrypt a message using the public key and sent it to the receiver. This person then uses the private key to decrypt the message [24]. RSA problem is given a positive integer n that is a product of two distinct odd primes p and q, a positive integer e such that gcd (e,(p-1)(q-1))=1, and an integer c, find an integer m such that me=c (mod n), while decryption is e . d=1 (mod (p-1). (q-1)). The security of the system depends on the difficulty of factoring the published divisor [18].

5. Conclusion

Data sharing concept can be defined as the process of interchanging, analyzing, retrieving and integrating data among multiple data sources in a controlled access manner. Although data sharing facilitates the way that data can be exchanged, security concerns arises a challenge for conducting data sharing, many polices include confidentiality and privacy must be taken into consideration. In this study we provide a literature review of security policies, focusing on privacy models that facilitate data sharing among different organizations in different areas. As a result for the study there are different data sharing model that applies different polices to preserve privacy, and semantic privacy preserving model overcomes many disadvantages of others models, and provide data integration, secure sharing through authorized view, in addition, each organization enable data sharing without affecting its clients.

6. Further Work

We are working on how to use the new Cryptosystem technique in proposing a new privacy preserving model depending on OneSwarm Model, then comparing the results after applying the new cryptosystem technique.

7. References

[1] Bakis, N., Aouad, G., Kagioglou, M., (2007), "Towards distributed product data sharing environments Progress so far and future challenges", *Elsevier-Automation in Construction*, 16, (5): 586-595.

[2] Clifton, C., Doan, A., Elmagarmid, A., (2004), "Privacy Preserving Data Integration and Sharing", *ACM- Research issues in data mining and knowledge discovery*: 19-26.

[3] Choi, J., Chun, S., Kim, D., Keromytis, A., (2013), "SecureGov: Secure Data Sharing for Government Services", *The Proceedings of the 14th Annual International Conference on Digital Government Research*.

[4] Freudiger, J., Rane, S., Brito, A., Uzun, E.**,** (2014),"Privacy Preserving Data Quality Assessment for High-Fidelity Data Sharing", *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Securit*, 21-29.

[5] Geambasu, R., Balazinska, M., Gribble, S., Levy, H., (2007), "HomeViews: Peer-to-Peer Middleware for Personal Data Sharing Applications", *ACM*: 235-246.

[6] Greif, I., Sarin, S., (1987), "Data Sharing in Group Work", *ACM*, 5, (2): 187-211.

[7] Harris, D., Khan, L., Paul, R., Thuraisingham, B., (2007), "Standards for secure data sharing across organizations", *ACM-Computer Standards and Interfaces*, 29,(1): 86-96.

[8] Hu, Y., Yang, J., (2011), "A Semantic Privacy-Preserving Model for Data Sharing and Integration", *ACM-Web Intelligence, Mining and Semantics*, (9): 1-12.

[9] Isdal, T., Piatek, M., Krishnamurthy, A., Anderson, T., (2010), "Privacy-Preserving P2P Data Sharing with OneSwarm", *ACM SIGCOMM Computer Communication Review*, 40, (4): 111-122.

[10] Mannai, D., Bugrara, K., (1993), "Enhancing Inter-Operability and Data Sharing In Medical Information Systems", *ACM*, 22, (2): 495-498.

[11] Sarathy, R., Muralidhar, K., (2004), "Secure and useful data sharing", *Elsevier*, 42, (1): 204– 220.

[12] Son, J., Kim, H., Kim, D., (2014), "On Secure Data Sharing in Cloud Environment", *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication*, 6.

[13] Varlamis, I., Vazirgiannis, M., (2001), "Bridging XML-Schema and relational databases A system for generating and manipulating relational databases using valid XML documents", *ACM*,: 105 - 114.

[14] Surendra, B., Manohar, T., (2014), "Advance Secure Multi-Owner Data Sharing for Dynamic Groups in the Untrusted Cloud", *International Journal of advanced technology and Innovative Research*, 6, (10): 1214-1219.

[15] Bauer, K., McCoy, D., Grunwald, D., Sicker, D., (2008), "BitBlender: Light-Weight Anonymity for BitTorrent", *ACM*.

[16] University if Miami, Confidentiality, Integrity and Availability (CIA). (2008). Retrieved April 26, 2015, from http://it.med.miami.edu/x904.xml.

[17] Rivest, R., Shamir, A., Adleman, L., (1978), "A Method for Obtaining Digital Signatures and Public-key Cryptosystems", Communications of the ACM, 21, (2): 120-126.

[18] BAUER, K., GRUNWALD, D., SICKER, D., (2009), "The Arms Race in P2P*", 37th Research Conference on Communication, Information and Internet Policy*.

[19] Fonville, M., (2010), "Confidential peer-to-peer file-sharing using social-network sites", *13th Twente Student Conference on IT*.

[20] Sasi, S., Dixon, D., Wilson, J., (2014), "A General Comparison of Symmetric and Asymmetric Cryptosystems for WSNs and an Overview of Location Based Encryption Technique for Improving Security", *IOSR Journal of Engineering*, 4, (3): 01-04.

[21] Rabin, M., (1979), "Digitalized Signatures and Public Key Functions as Intractable as Factorization", *Massachusetts Institute of Technology Laboratory for Computer Science*.

[22] Sharma, M., (2010), "Compression Using Huffman Coding", *IJCSNS International Journal of Computer Science and Network Security,* 10, (5): 133-141.

[23] Kakish, M., (2011), "Enhancing the Security of the RSA Cryptosystem", IJRRAS, 8, (2): 239-246.

