**Chapter 2**

**REVIEW OF LITERATURE**

There are not many methods proposed by researchers that help restrict or in that case prevent any fraud against the misuse of a patient’s private data. While a lot of research has been done on protecting an IT system against unauthorized access from attackers, not much work has been done on preventing sensitive information from being jeopardized, either due to abuse or carelessness, by authorized users. And since it is extremely critical that we protect sensitive data against any kind of malicious intent, there is an increasing need for a more advanced, and highly secure and accurate system.

There was a proposed system [1] which used a quantified-risk adaptive approach in order to keep a check on the access of a patient’s privacy. The paper claims that even though doctors are authorized to access their patient’s data, in reality, there is also an inherent risk in each access. Their proposed solution therefore, allows information consumers to have the freedom to choose what they want to access. A user’s data-accessing activities are associated with quantified risk scores, which will be added up over time. Request to access a resource is granted if doing so will not make the user’s aggregated risk exceed his/her tolerance threshold set by the system; otherwise, the request is denied. They do this by calculating a relevance-relation function. They determine what activity a certain honest doctor will perform with respect to a disease, hence calculating the probability that a certain record m will be accessed to serve purpose p. They also take special cases into consideration. Once the model is trained they expect that a malicious doctor would over access a patient’s data hence increasing the risk considerably.

The problem with this method is that it is not as accurate as one expects it to be. Another glaring flaw is that they have only considered one parameter to determine if a doctor is malicious, i.e. relevance of the data accessed. The solution thus ends up ignoring myraid of other factors that can contribute towards determining spiteful activities, thus proving to be inefficient.

In another paper [2], clustering algorithms were used to analyse and hence group hospital data for better management. The paper mainly focuses on order history. It processes these order histories in order to find out the temporal global characteristics of clinical activities. Once it has done so, it keeps applying clustering techniques to the results until they converge. The final output is then expected to be the optimum output. For example the output may be a particular day when there is highest activity for a particular test or the time of the day when there is least activity. The major drawback of a clustering based approach is that, since it is unsupervised, the clusters formed may be the opposite of ones we expect. For example if there were a lot of malicious overprescribed orders in the history database, those will form a cluster and hence end up as a category.

The comparison table of the studied reference papers is as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Reference Paper | Model | Drawbacks |
| [1] | Quantified Risk-Adaptive Access Control for Patient  Privacy Protection in Health Information Systems | A risk based adaptive model to keep a check on the data consumed by information consumers. | The system is less accurate and considers only one parameter i.e. relevance of data accessed. |
| [2] | Clustering-based Analysis in Hospital Information Systems | A clustering based algorithm to group orders based on various identified attributes | Clustering may lead to a lot of fake orders forming a cluster. |