INTRODUCTION

1.1. OVERVIEW:

The objective of data mining is to generalize across populations, rather than reveal information about individuals. The hitches that data mining works by evaluating individual data that subject to privacy concerns. Thus, the true problem is not data mining, but the way data mining is done. However, the concern among privacy advocates is well founded, as bringing data together to support data mining makes misuse easier. Much of this information has already been collected, however it is held by various organizations. Separation of control and individual safeguards prevent correlation of this information, providing acceptable privacy in practice. However, this separation also makes it difficult to use the information for purposes that would benefit to society, such as identifying criminal activity. Proposals toshare information across agencies, most recentlytocombatterrorism,wouldeliminatethesafeguardsimposedbyseparationoftheinformation.

1.2. MOTIVATION:

The title "Privacy Preservation in Collaborative Data Mining as Goal Oriented Attack Model" is justified by the implementation of Homomorphic Encryption to secure the mined data between the user and the server.

Thegoalofdataminingis toextractormineknowledgefromlargeamounts data. However, details often collected by several different sites. Privacy, legal and commercial concerns restrict centralized access to this data. Theoretical results from the area of secure multi party computation in cryptography prove that assuming the existence of trapdoor permutations; one may provide secure protocols for any two party's computationaswellasforanymultipartycomputationwithhonestmajority.However, the general methods are far too inefficient and impractical for computing complex

functions oninputsconsistingoflargesets ofdata.Whatremains openiscomeupwitha setoftechniques toachievethisefficientlywithinaquantifiablesecurityframework.The distributeddatamodel consideredistheheterogeneousdatabasescenariowithdifferent features ofthesamesetofdatabeingcollectedbydifferentsites.

This thesis argues thatitisindeedpossibletohaveefficientandpractical techniquesforusefulprivacy-preservingminingofknowledgefromlargeamountsof data. The dissertation presents several privacy preserving data mining algorithms operatingoververtICallypartitioneddata.Thesetof underlyingtechniquessolving independentsub-problemsarealsopresented.Together, theseenablethesecure“mining" ofknowledge.

In today’sinformation age,datacollectionisubiquitous,andevery transactionis recordedsomewhere.The resultingdatasets canconsistofterabytes orevenpetabytes of data,so efficiencyand scalabilityistheprimaryconsiderationofmost datamining algorithms.Datamining technology hasemergedasameansofidentifyingpatternsand trendsfromlargequantities ofdata.Mosttools operatebygatheringalldataintoacentral site,then runningan algorithm againstthatdata.However,privacy concernscan prevent buildingacentralizedwarehouseanddatamaybedistributedamongseveral custodians, noneof whichareallowedtotransfertheirdatatoanothersite.Theproblemisthat computingassociation rules.Thegoalistoproduceassociation rulesthathold globally while limitingthe information shared about each site. Previous workin privacy preservingdatamininghasaddressed twoissues.In one,the aimispreservingcustomer privacyby distortingthedatavalues. Theideaisthat thedistorteddatadoesnotreveal privateinformationandthusissafetouseformining.Thekey resultisthatthedistorted data,andinformationonthedistributionoftherandom datausedtodistortthedata,can beusedtogenerateanapproximationtotheoriginaldatavalues.

Recentadvancesindatacollection,datadissemination andrelated technologies haveinauguratedaneweraof researchwhereexistingdataminingalgorithmsshouldbe reconsideredfrom thepointofviewofprivacypreservation.Theneedforprivacyis sometimesduetolaw(e.g.,formedical databases)orcanbemotivatedbybusiness interests.However,therearesituationswherethesharingofdatacanleadtomutual