**CHAPTER TWO**

**LITERATURE REVIEW**

**2.1 Background of Health Information Privacy and Security**

Privacy is an underlying governing principle of the patient-physician relationship for effective delivery of healthcare. Patients are required to share information with their doctors to facilitate correct diagnosis and determination of treatment, especially to avoid adverse drug interactions. However patients may refuse to divulge important information in cases of health problems such as psychiatric behavior and HIV as their disclosure may lead to social stigma and discrimination (Applebaum, 2002). Over time, a patient‘s medical record accumulates significant personal information including identification, history of medical diagnosis, digital renderings of medical images, treatment received, medication history, dietary habits, and physicians‘ subjective assessments of personality and mental state among others (Mercuri, 2004).

Patient health information could serve a range of purposes apart from diagnosis and treatment provision. For example, information could be used to improve efficiency within health care system, drive public policy development and administration at state and federal level, and in the conduct of research to advance medical science (Hodge, 2003). Healthcare providers may use records to manage their operations, to assess service quality, and to identify quality improvement opportunities. Furthermore, providers may share health information through a regional health information organization to facilitate care services. Medical information of patients is also used for common good through federal and state government interventions regarding public health management, hospital accreditation, medical research, and for managing social and welfare systems.

**2.1.1 Threats to Cyber-health Community Security and Privacy**

Threats to patient-doctor interaction and information security could be categorized into two broad areas:

* Organizational threats that arise from inappropriate access of patient data by either internal agents abusing their privileges or external agents exploiting vulnerability of information systems.
* Systemic threats that arise from an agent in the information flow chain exploiting the disclosed data beyond its intended use (NRC, 1997).

**2.1.1.1 Organizational Threats**

This may assume different forms, such as an employee who accesses data without any legitimate need or an outside attacker (hacker) that infiltrates organization‘s information infrastructure to steal data or render it inoperable. At the outset, these organizational threats could be characterized by four components – motives, resources, accessibility, and technical capability (NRC, 1997). Depending on these components, different threats may pose different level of risk to organization requiring different mitigation and prevention strategies. Motives could be both of economic or noneconomic nature. For some, such as insurers, employers, and journalists, patient records may have economic value, while others may have noneconomic motives such as a person involved in romantic relationship. These attackers may have resources ranging from modest financial backing and computing skills to a well-funded infrastructure to threaten a patient as well as the operations of a healthcare organization. The attackers may require different types of access to carry out their exploits, such as access to the site, system authorization, and data authorization. In addition, threats could depend on technical capability of attackers who may be novice or sophisticated programmers. Moreover, with the growing underground cyber economy (Knapp and Boulton, 2006), an individual with the intent to acquire data and possessing adequate financial resources may be able to buy services of sophisticated hackers to breach healthcare data. Table 1 on the next page describes the various levels of access to a patient’s medical record by different categories of health care practitioners.

Recent studies suggest that the broad spectrum of organizational threats could be categorized into five levels, in the increasing order of sophistication (NRC and Rindfleisch, 1997):

* Accidental disclosure: healthcare personnel unintentionally disclose patient information to others, e.g. email message sent to wrong address or an information leak through peer-to-peer file sharing.
* Insider curiosity: an insider with data access privilege pries upon a patient‘s records out of curiosity or for their own purpose, e.g. a nurse accessing information about a fellow employee to determine possibility of sexually transmitted disease in colleague; or medical personnel accessing potentially embarrassing health information about a celebrity and transmitting to media.
* Data breach by insider: insiders who access patient information and transmit to outsiders for profit or taking revenge on patient.
* Data breach by outsider with physical intrusion:an outsider who enters the physical facility either by coercion or forced entry and gains access to system.
* Unauthorized intrusion of network system: an outsider, including former vengeful employees, patients, or hackers who intrude into organization‘s network system from outside and gain access to patient information or render the system inoperable.

Table 1: Access privileges in health care

|  |  |  |  |
| --- | --- | --- | --- |
| |  |  | | --- | --- | | **Level of Access** |  | | **Example** |
| **None** | **Outside attacker** |
| Site only | Maintenance worker |
| Site and System | |  | | --- | | Employee in the billing department who has access to information systems but not to clinical information | |  | |
| Data and System | |  | | --- | | Vendor or consultant with remote access privileges | |  | |
| Site, System, and Data | Care provider such as doctor or nurse |

**2.1.1.2 Systemic Threats**

(Etzioni, 1999)**,** in his work titled ‘*Limits to Privacy*‘, makes a strong case that a major threat to patient privacy occurs not from outside of the information flow chain in healthcare industry but from insiders who are legally privileged to access patient information. For example, insurance firms may deny life insurance to patients based on their medical conditions, or an employer having access to employees’ medical records may deny promotion, or worse, terminate employment. Patients and /or payer organizations may incur financial losses as a result of malpractices including up coding of diagnoses, and rendering of medically unnecessary services. In summary, healthcare information systems could be subjected to security threats from one or more sources including imposter agents, unauthorized use of resources, unauthorized disclosure of information, unauthorized alteration of resources, and unauthorized denial of service.

**2.1.2 Health Information Privacy Regulations**

In the last four decades, the US healthcare industry has undergone revolutionary changes, driven by advances in information technology and legislation such as the 1973 Health Maintenance Organizations Act. As personal health information is digitized, transmitted and mined for effective care provision, new forms of threat to patients‘privacy are becoming evident. In view of these emerging threats and the overarching goal of providing cost effective healthcare services to all citizens, several important federal regulations have been enacted including the Privacy and Security Rules under Health Insurance Portability and Accountability Act (1996) and State Alliance for eHealth (2007). The Privacy Rule of Health Insurance Portability and Accountability Act addresses the use and disclosure of a patient‘s protected health information by healthcare plans, medical providers, and clearinghouses, also referred as “covered entities”.

By virtue of their contact with patients, covered entities are the primary agents of capturing a patient‘s health information for a variety of purposes including treatment, payment, managing healthcare operations, medical research, and subcontracting. The Security Rule of HIPAA requires covered entities to ensure implementation of administrative safeguards in the form of policies and personnel, physical safeguards to information infrastructure, and technical safeguards to monitor and control intra and inter organizational information access.

**2.2 Privacy Concern among users of Cyber-Health Community**

A significant body of research has examined the perception of privacy concern from viewpoint of a special class of patients, including mental health patients, seekers of HIV testing, and adolescents using e-healthcare delivery platform. In a recent survey of past research on healthcare confidentiality, (Sankar, 2003) make four overarching conclusions. First, patients strongly believe that their information should be shared only with people involved in their care. Second, patients do identify with the need of information sharing among physicians, though HIV patients are less likely to approve sharing of their health information. Third, many patients who agree to information sharing among physicians reject the notion of releasing information to third parties including employers and family members. Lastly, the majority of patients who have undergone genetic testing believe that patients should bear the responsibility of revealing test results to at-risk family members.

This extensive body of research reviewed in (Sankar, 2003) mostly considered the use of identifiable or potentially identifiable information to other relevant entities including employers, families, and third parties. However, very limited research has examined patients‘ perception on sharing of anonymized cyber-healths, perhaps with exception of more recent studies that examine patients perception about consent to health information use for other than their own care (Bansal and Campbell, 2007). (Bansal, 2007), on the one hand, developed a set of constructs based on utility theory and prospect theory as antecedents of trust formation and privacy concern that impact users ‘personal disposition to disclose their health information to online health services websites.

In particular, this study reported that user‘s current health status, personality traits, culture, and prior experience with websites and online privacy invasions play a major role in user‘s trust in the health website and their degree of privacy concerns. On the other hand, (Cambell, 2007), in a mail based survey with adult patients in England, found that about 28% to 35% of patients are neutral to their health information – such as age, gender, ethnicity, reason for treatment, medical history, personal habits impacting health, type of treatment obtained, side effects of treatment – being used by physicians for other purpose. Only about 5–21% of patients, however, expected to be asked for permission to use their information by their physicians. Similarly only about 10% of the patients expected to be asked for permission if their doctors use their health information for a wide variety of purposes including, combining data with other patients’ data to provide better information to future patients, sharing how the treatment is working with other physicians in the hospital, teaching medical professionals, and writing research articles about diseases and treatments.

**2.3 Access Control in Cyber-health Platforms**

Modern healthcare related systems are large networked systems managing patient data with a multitude of users accessing health data for diverse contextual purposes within and across organizational boundaries. Role Based Access Control (RBAC), originally developed to manage access to resources in a large computer network (Ferraiolo and Kuhn 1992), is generally presented as an effective tool to manage data access in healthcare industry because of its ability to implement and manage a wide range of access control policies based on complex role hierarchies commonly found in healthcare organizations (Gallaher, 2002). This stream of research primarily focuses on developing algorithms and frameworks to facilitate role based information access (Motta and Furuie, 2003), and contextual access control (Covington, 2000). (Schwartmann, 2004) extends this stream of research by proposing an enhanced RBAC system that incorporates attributable roles and permissions. This enhanced system implementation is theorized to reduce the burden of managing access privileges by lowering extremely high number of permissions and roles to a manageable size and hence reducing administrative cost. In addition progress is being made in several fronts, including use of autonomous agents to create privacy-aware healthcare applications (Tentori, 2006), authorization policy framework for peer-to-peer technology based distributed healthcare system (Al-nayadi and Abawajy, 2007), encrypted bar code technology framework for electronic transfer of prescription (Ball et al. 2003), pseudonymous linkage (Reidl, 2007), and electronic consent models that allows patients to define which component of a medical record could be shared to whom (O‘Keefe and Nepal, 2006).

**2.4 Information Security for Authorized Data Disclosure**

In healthcare sector, it is often necessary to share data across organizational boundaries to support the larger interests of multiple stakeholders as well as agencies involved with public health. However, the release of patient‘s data could entail personally identifying information as well sensitive information that may violate privacy as well cause socioeconomic repercussions for patient. Yet, such data, when masked for identifying and sensitive information, must maintain the analytic properties to assure statistical inferences, especially when released for epidemiological research (Truta, 2004). Advances in technology have led to consolidation of cyber-healths from multiple sources to a single research database which supports researchers engaged in improvement of public health, clinical methods and health services in general.

A significant and growing body of research, building on the theory of statistical disclosure control , offers a diverse range of data masking methodologies and frameworks to minimize or control the disclosure risk of patient information e.g. global and local recoding (Samarati, 2001), micro-aggregation (Domingo-Ferrer and Mateo-Sanz, 2002), data perturbation (Muralidhar and Sarathy, 2005), data swapping (Dalenius and Reiss, 1982), and data encryption (Chao, 2005) among others, in addition to de-identification or removal of data identifiers (Ohno-Machado, 2004).

**2.5 Security Protocols in Cyber-health System**

Thereare some ethical priorities of cyber-healthcare system which include the following:

**2.3.1 Patient’s Healthcare Information Security**

The National Institute of Standards and Technology (NIST), the federal agency responsible for developing information security guidelines, define security as the preservation of data confidentiality, integrity, availability (commonly referred to as “CIA” triad). Security can be defined as “procedural and technical measures required to prevent unauthorized access, modification, use and dissemination of data stored or processed in a computer system and to protect the system in its entirety from physical harm” (Ware, 1976). Security help keep cyber-health safe from unauthorized use. When someone hacks into a computer system, there is a breach of security (and also potentially, a breach of confidentiality). No security measure, however, can prevent invasion of privacy by those who have authority to access the record (Gostin, 1995). Security also refer to maintaining integrity of electronic medical information and ensuring availability to those who need access and are authorized to view such clinical data, including images, for the purpose of health care. The increasing concern over the security of health information stems from the rise of electronic cyber-healths, increased use of mobile devices such as smartphones, medical identity theft, and the widely anticipated exchange of data between and among organizations, clinicians, and patients. If patient trust is undermined, they may not be forthright with the physician. For the patient to trust the doctor, records in the office must be protected. Medical staff must be aware of the security measures needed to protect their patient data and the data within the practices.

The responsibility that the doctor have to protect their patient from harm extends to protecting patient information, privacy and confidentiality. Patient information security includes the steps healthcare providers must take to guard patient “protected health information” commonly referred to as PHI, from unauthorized access or breaches of privacy or confidentiality. A recent survey found that 73 percent (73%) of physicians text other doctors about work to know if all these devices are encrypted or not.

**2.3.2 Patient Privacy**

Patient privacy refers to the right of patients to determine when, how and to what extent their health information is shared with others. It involves maintaining confidentiality and sharing identifying data, known as protected health information (PHI), only with healthcare providers and related professionals who need it in order to care for the patient. However, although privacy is often used interchangeably with the terms “confidentiality” and “security,” they have distinct meanings.Privacy addresses the question of who has access to personal information and under what conditions. Privacy is concerned with the collection, storage, and use of personal information, and examines whether data can be collected in the first place, as well as the justifications, if any, under which data collected for one purpose can be used for another (secondary) purpose. An important issue in privacy analysis is whether the individual has authorized particular uses of his or her personal information (westin, 1967).

Justine Warren and Brandeis define privacy as the right “to be let alone”. According to Richard Rognehaugh, it is “the right of individuals to keep information about themselves from being disclosed t others; the claim of individuals to be let alone, from surveillance or interference from other individuals, organizations or the government”. The information that is share as a result of clinical relationship is considered confidential and must be protected. The various forms (including identification data, diagnoses, treatment and progress notes, and laboratory results) and can be stored in multiple media (e.g. paper, video, electronic files). Information from which the identity of the patient cannot be ascertained. For example, the number of patient with prostate cancer, number of patients with diabetics, number of patient with asthma and so on.

Patient information should be released to others only with the patient permission. This is not, however, to say that the doctors cannot gain access to patient information. Information can be released for treatment, payment, or administrative purposes without a patient’s authorization.

**2.3.3 Confidentiality**

Confidentiality may seem a very straight forward principle, but translating principle into practice can be problematic. There are all sorts of situations where it is difficult to know if patient information should be shared or not. The duty of confidentiality goes beyond undertaking not to divulge confidential information. It include a responsibility to make sure that written patient information is kept securely. Confidential records should not be left where other people may have access to them and information about patients should be sent under private and confidential cover, with appropriate measures to ensure that it does not go astray. Patient should be informed about the kind of information being held about them, how and why it might be shared, and with whom it might be shared. Patient information leaflets are a convenient way of notifying patients about this, nut they are not sufficient in themselves. Bear in mind that few patients will bother to read the leaflets, and some may not be able to read them. Confidentiality is not an absolute principle, and there are circumstances where it is permissible to disclose a patient’s medical records to a third party. Confidentiality safeguards information that is gathered in the context of an intimate relationship. It addresses the issue of how to keep information exchanged in that relationship from being disclosed to third parties (Westin, 1976). Confidentiality, for example, prevents physicians from disclosing information shared with them by a patient in the course of a physician–patient relationship. Unauthorized or inadvertent disclosures of data gained as part of an intimate relationship are breaches of confidentiality (Gostin and Hodge, 2002).

**2.3.4 Integrity and Availability**

Integrity assures that the data is accurate and has not been changed. This is a broad term for an important concept in the electronic environment because data exchange between systems is becoming common in the health care industry. Data may be collected and used in many systems throughout an organization and across the continuum of care in ambulatory practices, hospitals, rehabilitation centers, and so forth. This data can be manipulated intentionally or unintentionally as it moves between and among systems. It is essential to maintain data integrity when transferring information by verifying that the information arrived as it was sent and was not modified in any way. Methods to maintain data integrity include intrusion detection such as tripwire, and message digest or hashing to detect any alteration of the data.

Poor data integrity can also result from documentation errors, or poor documentation integrity. A simple example of poor documentation integrity occurs when a pulse of 74 is unintentionally recorded as 47. Whereas there is virtually no way to identify this error in a manual system, the electronic cyber-health has tools in place to alert the clinician that an abnormal result was entered.

Features of the cyber-health system should not allow data integrity to be compromised. Take, for example, the ability to copy and paste, or “clone,” content easily from one progress note to another. This practice saves time but is unacceptable because it increases risk for patients and liability for clinicians and organizations. Clinicians and vendors have been working to resolve software problems such as screen design and drop-down menus to make electronic healthcare systems both user-friendly and accurate.

The availability of information is that, if the system is hacked or becomes overloaded with requests, the information may become unusable. To ensure availability, cyber-health systems often have redundant components, known as fault-tolerance systems, so if one component fails or is experiencing problems the system will switch to a backup component.

**2.3.5 Authentication**

In addition to the importance of security, patient privacy and integrity and availability, the electronic healthcare system must address the authentication of authorized users. Authentication is the process of verifying the identity of a user to a computer system and can be accomplished using login passwords, digital certificates, smart cards and biometrics.

Authentication only verifies the identity of an individual. It does not define their access (authorization) rights. Authentication is the act of confirming the truth of an attribute of a single piece of paper claimed true by an entity. In contrast with identification which refers to the act of stating or otherwise indicating a claim purportedly attesting to a person or things’ identity, authentication is the process of actually confirming that identity. It might involve confirming the identity of a person by validating their identity documents, verifying the validity of a website with a digital certificate, tracing the age of an artifact by carbon dating or ensuring that a product is what its packaging and labelling claim to be. In other words, authentication often involve verifying the validity of at least one form of identification.

**2.5.4 Ensuring Security and Privacy in Cyber-health Community using Trust Management**

Cyber-health usage can provide superior security and privacy solutions once the healthcare system is implemented. The cyber-health system enhance administrative controls by guiding employees through the stringent privacy and security training process. It also help to monitor physical and system access by creating physically inaccessible systems to unauthorized individuals. Cyber-health system also monitor system user by detecting any security breach or attempt at a breach. Trust management should also be applied by disguising all data through cryptography. Reliable cyber-health community service providers apply these enhanced security and privacy protocols. Perhaps the most important security protocol is data encryption, which causes data to become unreadable to outside sources.

**2.6 Overview of Trust in Information Technology**

These concepts have heightened relevance in the last decade in computer science, particularly in the area of distributed artificial intelligence. The multi-agent system paradigm and the growth of e-commerce have increased interest in trust and reputation. In fact, trust and reputation systems have been recognized as the key factors for electronic commerce. These systems are used by intelligent software agents as an incentive in decision-making, when deciding whether or not to honor contracts, and as a mechanism to search trustworthy exchange partners. In particular, reputation is used in electronic markets as a trust-enforcing mechanism or as a method to avoid cheaters and frauds. (Dellarocas, 2003)

Another area of application of these concepts in agent technology, is teamwork and cooperation. (Montaner et al., 2002). Trust and reputation both have a social value. When someone is trustworthy, that person may be expected to perform in a beneficial or at least not in a suspicious way that assure others, with high probability, good collaborations with him. On the contrary, when someone appears not to be trustworthy, others refrain from collaborating since there is a lower level of probability that these collaborations will be successful. (Gambetta, 2007)

Trust is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently or his capacity ever to be able to monitor it) and in a context in which it affects his own action.

**2.7 Computational Trust**

In information security, computational trust is the generation of trusted authorities or user trust through cryptography. In centralized systems, security is typically based on the authenticated identity of external parties. Rigid authentication mechanisms, such as Public Key Infrastructures (PKIs)or Kerberos have allowed this model to be extended to distributed systems within a few closely collaborating domains or within a single administrative domain. During recent years, computer science has moved from centralized systems to distributed computing. (Weise, 2001; Kohl and Neumann, 1993). This evolution has several implications for security models, policies and mechanisms needed to protect users’ information and resources in an increasingly interconnected computing infrastructure. (Seigneur, 2005).

Identity-based security mechanisms cannot authorize an operation without authenticating the claiming entity. This means that no interaction can occur unless both parties are known by their authentication frameworks. Spontaneous interactions would, therefore, require a single, or a few trusted certificate authorities (CAs). In the present context, PKI has not been considered since they have issues, thus it is unlikely that they will establish themselves as a reference standard in the near future. A user who wishes to collaborate with another party can choose between enabling security and thereby disabling spontaneous collaboration, or disabling security and enabling spontaneous collaboration. It is fundamental that mobile users and devices can authenticate in an autonomous way without relying on a common authentication infrastructure. In order to face this problem, we need to examine the challenges introduced by "global computing",a term coined by the EU for the future of the global information society, and to identify their impact on security.

The term ―trust management‖ was coined by Matt Blaze of AT&T Research Labs, in 1996 (Blaze et al., 1996). Blaze’s concept of trust management centered on specifying security policies and applying them to authorization statements embedded in credentials to enable a trust management engine to directly assess whether a requested action should be allowed. Blaze’s use of the term reflects trust based on policy enforcement. More recently Tyrone Grandison (Grandison, 2003) defined trust more generally to include both trust based on policy enforcement and trust based on a good reputation. He defines trust management as ―the activity of collecting, encoding, analyzing and presenting evidence relating to competence, honesty, security or dependability with the purpose of making assessments and decisions regarding trust relationships‖ (Grandison, 2003).

Trust management systems are generally based on reputation evidence (Lin et al., 2005), credential-based evidence supporting policy enforcement (Seamons et al., 2003) or both (Grandison, 2003). Trust management is most commonly used in distributed systems where there is no single authority, such as an employer or civil authority, that can control and levy punishment for inappropriate actions, and for which blind trust is not a viable option (e.g., because of the probability of a trust violation coupled with the associated risks).

Reputation-based trust management mechanisms are common in e-commerce (Carman et al,.2000) where a consumer needs to minimize the risk involved in using a service (e.g., purchasing a product) from an unfamiliar provider. Reputation is also useful in electronic communities, such as peer-to-peer (Xiong and Liu. 2004) or wireless sensor networks to detect nodes that are not being good citizens of the community, such as entities that freely consume resources without offering any resources in return. The risk to be avoided may be maliciousness or simply poor QoS. Trust management was originally conceived as a method for establishing trust across security enclaves where authentication is either impossible or meaningless, but reputation-based trust is also useful for maintaining trust within a security enclave when insider threat, intruders, or deteriorating QoS is a concern.

Reputation-based trust management works best when the principals have:

* Defined interactions with other principals over a period of time,
* Measureable elements of satisfactory and unsatisfactory behavior,
* A large community of peers to contribute trust observations, and
* A large number of transactions with peers.

Credential-based trust management systems typically have to do with authorization or delegation. In credential-based trust management systems, formally specified policy statements are used by service providers or resource owners in conjunction with authentication certificates and/or authorization credentials to determine if consumers have the right to use a service or resource. This process may occur through a trust negotiation process wherein each entity iteratively reveals either policy statements or credentials to the other until trust has been established. Although authentication certificates, such as X.509 certificates, are useful within a security enclave to prove identity, attribute credentials and peer-granted identity certificates (e.g. PGP (Pirzada and McDonald, 2004)) can be used to control access to resources in distributed systems that cross security enclaves where there are no shared authentication certificates controlled by a central authority. Even when a central authority spans enclaves, authentication merely serves to prove identity, not trustworthiness.

Credentials also provide much more expressiveness than the typical read/write/execute access control permissions. For example, an authorization credential might represent that ―the holder is authorized to sign contracts worth up to $200,000‖ or a policy may require that only university students are eligible to sign up for a benefit and therefore a ―student‖ attribute credential signed by a known university must be supplied.

Credential-based trust management works best when the principals have:

* Well-defined activities and interactions
* A shared trust anchor

Both forms of trust management have the potential to provide value to swarmbased autonomic computing systems.

**2.8 Review of Related Works**

Trust and reputation have gained importance in diverse fields such as economics, evolutionary biology, distributed artificial intelligence, grid computing, agent technology, among others. Various definitions for trust and reputation have evolved as a result; in what follows, we review the work done on trust and reputation in the specific field of agent technology.

Mui et al. propose in (Mui et al., 2002) a computational model based on reciprocity (a mutual exchange of deeds, either in favor or revenge), trust (a subjective expectation an agent has about another’s future behavior based on the history of their encounters) and reputation (a perception that an agent creates through past actions about its intentions and norms). Mui et al. also propose in (Mui et al,. 2003) an intuitive typology summarizing different notions of reputation that have been studied across diverse disciplines. The typology divides reputation into several components: individual reputation which includes direct and indirect reputation, and group reputation. Direct reputation includes interaction-derived reputation and observed reputation, whereas indirect reputation includes prior-derived reputation, group-derived reputation and propagated reputation. The relative strengths of these different notions of reputation were studied in a set of evolutionary games.

Abdul-Rahman et al. (Abdul-Rahman and Hailes, 2001) study reputation as a form of social control in the context of trust propagation — reputation is used to influence agents to cooperate for fear of gaining bad reputation. Although not explicitly described, they have considered reputation as a propagated notion which is passed to other agents by means of “word-of-mouth”.

Pujol et al. (Pujol et al., 2002) propose a method of calculating reputation based on the position of each member of a community within the corresponding social network. A new algorithm, NodeRanking, was developed to obtain such measures. This algorithm assesses reputation by using local information only.

Cubaleska and Schneider (Cubaleska and Scheider, 2002) propose a method for a posteriori identification of malicious hosts to build a trust policy. Depending on how much the source host trusts the other hosts, it can either define an appropriate order in which selected hosts should be visited, or it can decide which hosts it does not want to contact again.

Barber and Kim (Barber and Kim, 2002) propose a computational model that combines belief revision and trust reasoning processes and show how deceptive or incompetent agents can be isolated from an agent’s decision making process with this model. An agent learns reputations of other agents using dissimilarity measures calculated from the previous belief revision processes (Direct Trust Revision) and/or communicated trust information that contains reputations (Recommended Trust Revision).

Agents utilize this model to detect fraudulent information and to identify potential deceptive agents as a form of social control in which an individual member is responsible for taking care of security.

In another work, Barber, et al. (Falcone et al., 2003) discuss the difficulty of assigning initial reputations when either the truster or the trustee is new to the system. They state that some interaction must take place for recommendation– based reputations to build up, and that when initial reputation assignments are arbitrary or have default values, a stable base for reputation has not yet been reached. Since information collection takes time, agents have to evaluate the time available for decision-making against the increased reputation base stability gained as more reputation processing is performed. Furthermore, Barber and Kim (Barber and Kim, 2000) define trust as the agent’s confidence in the ability and intention of an information source to deliver correct information. Reputation on the other hand, is defined as the amount of trust an information source has created for itself through interactions with other agents. They also propose a multi-agent belief revision algorithm that utilizes knowledge about the reliability or trustworthiness of information sources.

Jurca and Faltings (Jurca and Faltings, 2002) state that the most reliable reputation information can be derived from an agent’s own experience. However, much more data becomes available when reputation information is shared within an agent community. They attempt to encourage agents to truthfully share reputation information by providing incentives (a side-payment scheme) for recommenders to tell the truth. They point out that it is not in the best interest of an agent to truthfully report reputation information because reporting reputation information provides a competitive advantage to others and by reporting positive ratings an agent slightly decreases its own reputation with respect to the average of other agents.

Tran and Cohen (Tran and Cohen, 2003) propose a reputation-oriented reinforcement learning algorithm for buying agents in electronic market environments, taking into account the fact that the quality of a good offered by different selling agents may not be the same and that a selling agent may alter the quality of its goods. Modeling the reputation of sellers allows buying agents to focus on those sellers with whom a certain degree of trust has been established. The authors also include the ability for buying agents to explore the marketplace in order to discover new reputable sellers.

Finally, the work done by Azzedin et al. (Azzedin and Maheswaran, 2002) in the field of trust and reputation in grid computing systems. They present a formal definition of behavior trust and reputation and discuss a behavior trust management architecture that models the process of evolving and managing behavior in grid computing systems.