INDIAN STATISTICAL INSTITUTE

KOLKATA

DISSERTATION

TRUST BASED EVENT MODEL

A dissertation submitted in partial fulfillment of the requirements for the award of Master of Technology in Computer Science

by

Sachu Thomas Isaac

MTC1214

Under the supervision of **Prof. Aditya Bagchi**,

Electronics and Communication Sciences Unit, Kolkata

CERTIFICATE

This is to certify that the thesis titled **Trust based Event Model** submitted by **Sachu Thomas Isaac** is in partial fulfilment for the award of the degree of Master of Technology is a bona-fide record of work carried out by him under my supervision. The thesis has fulfilled all the requirements as per the regulations of this Institute and, in my opinion, has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other university for the award of any degree or diploma.

Aditya Bagchi
Professor
Electronics and Communication Sciences Unit,
Indian Statistical Institute, Kolkata

Date: 11th July, 2014

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my supervisor, Prof. Aditya Bagchi, for providing me with this chance to work under his guidance. I share my heart felt gratitude to him for providing me with an adequate problem statement which was the foundation to this dissertation work.

I also like to thank all the professors of the Indian Statistical Institute for all the support and guidance they provided me during the 2 years course time of M.Tech. Computer Science.

I would like to thank all my classmates and seniors for their morale and relentless support for making this journey smooth and worth it. I would like to thank my parents and sister for being the driving force they are in my life.

ABSTRACT

Over the last decade, information generation and distribution has gone through a complete over haul. Information generation is no more captured in the hands of few power centres and has expanded into a social platform. This growth has been primarily spear headed by citizen reporting, smart mobile devices and social media where information are generated by the people, for the people and about the people. This dissertation work attempts to model such a scenario where the temporal factor and the trustworthiness of such information is of utmost importance.

CONTENTS

Chapter 1

1 INTRODUCTION

Why do we need events?

Over the years, information technology has generated enormous amount of information that are interrelated and time stamped. Each such information may be of any form, multimedia or text. Researchers are continually developing new models to capture, store, organize and query such information. Such information, in short, may be called events. An event in a simplistic way is something that has happened, is happening or is expected to happen.

In the early stages of development of event model, researchers considered event as a change in the database when an operation is performed on it. This can be user triggered or system triggered (exceptions). The complexity of events were high, as the event is generated from multimedia sources or streaming data. Till now, the systems did not consider an event as a fundamental information unit which can be stored, queried and merged with other events. Later on, researchers have developed systems where each data object is coupled with a locational and temporal attribute.

How are events generated?

Events can be generated by users, who give explicit information related to themselves or about the environment they are in. Events can also be generated from information retrieval systems which extract information from a known source. Researchers are continually working towards a robust system that can extract information from a data stream. The data stream may be a video, audio or even a log file.

The focus

Event models developed so far, consider any source to produce accurate events from information ecosystems. In other words, we completely trust the source for the authenticity of such events. The present research effort tries to focus on the authenticity of the events generated from various sources. How do one know which events are genuine or partially incorrect or completely distrustful. How trustworthy is the source?

This dissertation proposes a model which has a trust layer over the existing event model and tries to quantify the trust in such a way that the events can be compared.

Chapter 2

2 SURVEY

2.1 EVENT MODEL

Here, we assume a model which consists of various sources which generate events and each event generated are stored and queried. More formally, an event is associated with at least one data object that is in a state for some finite amount of time or it undergoes a change in state. Researchers have tackled the problem of efficiently storing, organising and querying the event model. Some of the traditional approaches are:

Active Databases

In traditional databases, only the queries and updates are supported but active databases are reactive (Díaz, 1999), that is take an action if it encounters an event. An event in such a case can vary from a structure operation on the database, user defined or system generated.

Complex Event Processing

This system (C.Luckham, 2001) considers a distributed world where events occur and notify some event handlers who process the event and may generate its own notifications. It not only mediates the information in form of events between providers and consumers, but support the detection of relationships among events.

Event-oriented spatiotemporal databases

This type considers every data objects to be coupled with a locational attribute and a temporal attribute that record the spatial and the valid time extent associated with the object (Duan, 1995). So each spatial-temporal object associates to events as a data organising attribute. In this case event is interpreted as a change in data property.

E^* - A Graph-based event model using RDF and Ontologies

In the previous models, events were viewed as changes in a value (relational model) and transitions of class memberships, or participation in relationships. In real life systems, the number of relationships cannot be known at the design time. This drives for a new model which can take into modelling of relationships and its properties on the fly. Researchers have suggested many models that use Resource Descriptive Format (RDF), a World Wide Web standard for the semantic web, as the primary formal structure for representing and querying over graphs.

One of such solutions using Resource Descriptive Format (RDF) and graph structure in event model is E^* model (Jain A. G., 2011). This is an extension of the E event model as suggested by (Jain U. W., 2007). Graph based representations are suitable for cases where the number of relationships between data objects is very large, graph-traversal is important for query evaluation, and graph properties can be used to query, infer and analyse data. In addition, many of these groups consider a knowledge model for describing and querying events and adopt the Web Ontology Language (OWL) as the formal knowledge representation model which is also graph structured, and lends itself to a limited form of logical inference.

A major difference between this ontological model with the previous spatiotemporal data models is that states, events and elements of the event are directly represented as first class model constructs and not inferred entities derived from changes in value or class membership.

2.2 TRUST BASED ACCESS CONTROL

Various trust based access control (TBAC) models have been developed over years.

TBAC for a peer to peer system – Approach 1

In peer to peer (P2P) file sharing system, the task of controlling access to sharing information is more difficult due to the decentralised and anonymous characteristics of the P2P systems. In this model the authors, Huu Tran et al (Huu Tran, 2005) have proposed an access control framework based on the discretionary access model. This leaves the control of access rights to the discretion of the owner of the object or file. Here, each file is assigned with two thresholds which capture the aspect of size and content. The access values are relative and assessed on a P2P basis. The computation of trust is done from combinations of four different scores : direct trust, indirect trust, direct contribution, and indirect contribution.

DIRECT TRUST represents the host's belief on the client's capacities, honesty and reliability based on the host's direct experiences. The quantification of this type of trust is done by Beth et al' formula [*paper reference*] as

$$T_{ij} = 1 - \alpha^{n}$$

where T_{ij} denotes the trust value that peer *i* has in peer *j*,

n is the number of satisfied transactions with peer j

and α is the learning rate in the interval of [0,1].

INDIRECT TRUST represents the host's belief on the client's capacities, honesty and reliability based on the recommendations from other peers. Consider T_{it} as the direct trust of peer *i* has on peer *t* and T_{tj} as the direct trust of peer *t* has on peer *j*, the indirect trust of peer *i* has on peer *j* is given by

$$R_{ij} = \left(\sum_{t=1}^{k} T_{it} * T_{tj} \right) \Big/ k$$

DIRECT CONTRIBUTION measures the contribution of the client to the host in term of information volume downloaded and uploaded between them. Direct contributions is measured in megabytes. It indicates the relative transferring volume of shared information from the client peer to the host peer over their interaction history. Hence, direct contribution is defined as

$$Q_{ij} = D_{ij} - D_{ji}$$

where D_{ij} denotes the amount of information that peer *i* has downloaded from peer j, similarly D_{ji} denotes the amount of information that peer *j* has downloaded from peer *i*.

INDIRECT CONTRIBUTION measures the contribution of the client to the network in terms of information volume the client exchange with other peers. The indirect contribution is evaluated based on the recommendations from different peers which needs to weighted differently, depending on the trust level on the recommending peer. The indirect contribution is as follows,

$$P_{ij} = \sum_{t=1}^{k} T_{it} * \, Q_{tj}$$

where, Q_{tj} is the direct contribution score of peer t to peer j and T_{it} is the direct trust of peer *i* on peer t.

As mentioned earlier, this system uses a two level threshold for trust and contribution. The overall trust value is a weighted summation of direct trust and indirect trust. The overall contribution score is a weighted summation of direct contribution and indirect contribution. In principle the host sets the threshold for both of these quantities for every file, and the client will download the files only if its values exceeds the threshold of the host's file.

TBAC for peer to peer system – Approach 2

Bin Yu et al (Bin Yu, 2004) suggested a reputation based mechanism. In order to evaluate the trustworthiness of a peer, the peers must rely on incorporating the knowledge of other peers who have interacted with the same peer. The research group considers ratings as the trust value. They have defined two types of rating depending on the prior interactions with the peer.

LOCAL RATING is based on the direct interactions with the second peer. The rating is generated every time when an interaction takes place. Suppose peer P_i has rated the quality of service of the latest h interactions with P_j as a series of probabilistic ratings, $S_{ij} = \{ s_{ij}^1, s_{ij}^2, s_{ij}^3, \dots, s_{ij}^h \}$ where $0 \le s_{ij}^k \le 1$, and h is bounded by the allowed history H. The local rating or the reliability of peer P_i for P_j can be computed in the following two ways:

1. Simple averaging

$$R\left(P_i, P_j\right) = \begin{cases} \sum_{k=1}^{h} s_{ij}/h & h \neq 0 \\ 0 & h = 0 \end{cases}$$

2. Exponential averaging

$$R\left(P_i, P_j\right) = \begin{cases} \gamma \, [\, s_{ij}^h + \cdots + (1-\gamma)^h \;\, s_{ij}^1 \,] & h \neq 0 \\ 0 & h = 0 \end{cases}$$

Where γ $(0 \le \gamma \le 1)$ is the averaging constant and determines the weights given to the most recent past observations. The bigger the γ is, the faster the past observation is forgotten.

AGGREGATE RATING combines the local ratings with testimonies received from other peers. Suppose $\{W_1, W_{2,} \ldots, W_L\}$ are a group of peers who incorporate knowledge of peer P_j and the testimony $R\left(W_k, P_j\right)$ is witness W_k's local rating for P_j, w_k is the weight for the credibility of witness W_k, then the prediction from the testimonies is

$$\mathcal{P} = \begin{cases} \sum_{k=1}^{L} w_k * R\left(W_k, P_j\right)/L & L \neq 0 \\ 0.5 & L = 0 \end{cases}$$

The aggregate rating towards peer P_j is

$$T\left(P_i, P_j\right) = \begin{cases} \eta R\left(P_i, P_j\right) + (1 - \eta)\, \mathcal{P} & L \neq 0 \\ 0.5 & L = 0 \end{cases}$$

where η is peer $P_i's$ confidence about its local rating for peer P_j and $\eta = h/H$; L is the number of peers found by P_i and $1 \leq k \leq L$. If the aggregate rating form testimonies is above a threshold then peer P_i will interact with P_j.

Chapter 3

3 Fusion of trust model and event model

Information systems usually model a perspective of the real world scenario. Early requirements show that a relational database would suffice the storing and querying of data where the data is treated as objects and attributes are defined along with them. But with the advances in computing and communication technologies, the need to restructure the approach towards a more flexible and tangible models is necessitated. We see an endless stream of structured, semi-structured or unstructured data that needs to be stored and queried upon. One such type of data are events, where in the modern age holds a substantial importance. How each events are related? What can be inferred from a set of events? In an abstract and more generalized setting, can we trust the source which generated the event? Thus, a trust layer is necessary and viable to determine and help to answer such questions.

3.1 The setting

In our model, we have different sources which are capable of generating any type of event. These sources may be information retrieval systems themselves or a user or a logging system which generates a continuous stream of time coded information or a multimedia system. Each source has an additional attribute called the *source_trust* which quantifies how trustworthy the source is.

A category called verifiers is infused in the model which is nothing but a set of sources which have a 'high' trust value. The word 'high' is very qualitative but in an actual application we can assign a threshold value and sources whose *source_trust* value is greater than that threshold value are categorized as verifiers. Each event generated from a source has an attribute called the *event_trust*. This attribute defines the authenticity of the event.

3.2 The model

The attributes *source_trust* and *event_trust* has a value between 0 and 1. In the initial setting, all *source_trust* values are given a value of 0.5. When an event is generated from a source, the initial value of the *event_trust* is the same as the *source_trust*. The consumer of the event or the system needs to authenticate the event information which was generated.

The system will send the event to the known verifiers which are closest to them for authentication. The number of such verifiers in the vicinity of the consumer may vary largely. Let the *source_trust* attribute be denoted as S_{ij} and the *event_trust* attribute be denoted as E_j where *i* denotes the source id and j denotes the event id.

Before verification phase :

$$E_j = S_{ij} \text{ (initial setup)}$$

When the event is sent for verification to sources with high *source_trust* value (crossing an assumed threshold) we need to automate the update changes in the trust values. There are two key factors in this phase. Firstly, when the system or consumer of events verifies from a number of sources, how will the *event_trust* value be affected. Secondly, the *source_trust* value of the source which generated the event should also be changed depending on the response of the verifiers. These updates are effected with the following equation.

$$E_j = \omega * S_{ij} + (1-\omega) * (1-\alpha^n)$$

where, ω is the weight and $0 \leq \omega \leq 1$,

α is the learning rate and $0 \leq \alpha \leq 1$

N is the number of verifiers who gave positive response.

The *event_trust* value is updated for a specific values of (ω_1, α_1) and the *source_trust* value is updated with another set of values (ω_2, α_2) depending in application domain.

The physical meaning of the quantities (ω, α, n)

The weight ω interprets on how much weightage needs to be given to the previous value of the *source_trust* in other words it correlates to importance of the information. The weight ω will be comparatively lesser when used in the computation of final *source_trust* than when used in the computation of final *event_trust*.

The learning factor, α corresponds to how quickly should the value increase to 1 as n increases. This factor is crucial as it solely corresponds to the change in trust value. The figure below shows the plot for the function $(1-\alpha^n)$ for different values of α.

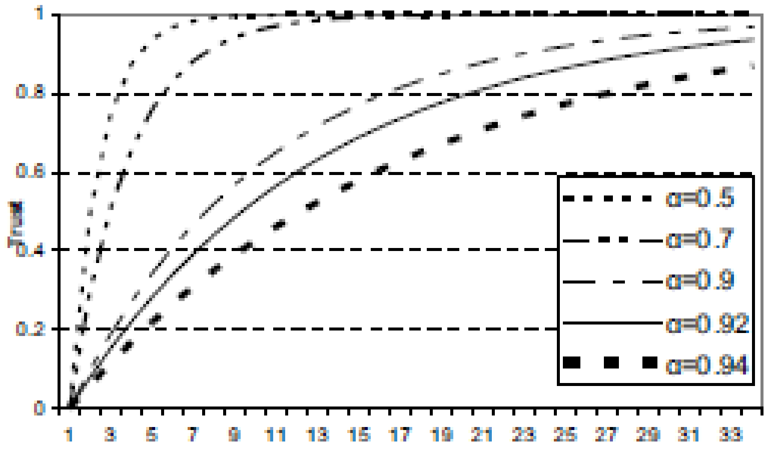

Figure 1: Trust values against the number of positive verifications

The number of verifiers, N corresponds to the number of positive responses for that particular event. Each source which are assigned the role of a verifier, has a set of responses { YES, NO, CANTSAY }.

3.2.1 Example

An agent of Newsan went to TroubleTown, 60 miles east of StateCapital on September 12, 2010 to cover a rally by EthnicMinority, a group that called the rally to protest against the noReservation law passed by theGovernment three days back.The rally started at 2 pm.The main speaker EthnicLeader made some inciting comments about how the Government must be stopped from doing its regular business unless the job reservation demands of the group were met.

Twenty minutes into the speech, a section of the crowd grew violent and started throwing stones at city buses. Soon the violence spread, and within the next half hour, the mob set fire to a police vehicle, damaged a fire truck and some private vehicles. The police immediately started firing in the air to disperse the crowd and called for additional forces. The violence was brought under control an hour after the additional forces arrived. Later, EthnicMinority reported that several rally-goers were injured from the clash with the police.

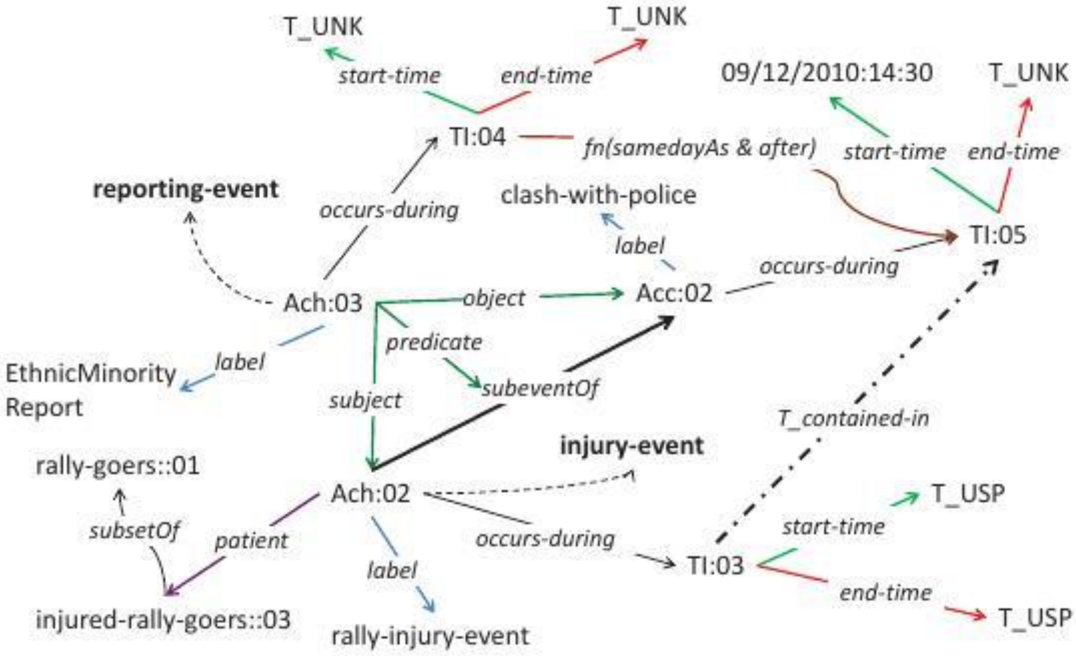

Figure 2 : An instance of an event in the example represented in the form of triples.

3.2.2 Implementation

In the above example, with the assumption that an information retrieval system extracted the various events pertaining to the Rally and the distress caused due to the rally. Each such event is modelled as a set of triples (subject, predicate, object) and stored as Resource Descriptive Framework (RDF) statements. Each such set consists of the various properties associated with an event including the occurs_during and occurs_at attribute relating to the spatiotemporal details and the *event_trust* attribute.

A general structure of each event is

```
type event::subClass*(DOLCE::perdurant)
        ( event_id integer,
          occurs_during timeInterval,
          occurs_at Location multiple optional,
          observed_by union(Person, ImageDevice) multiple optional,
          experienced_with Media multiple optional )
```

Each such event generated from a source, like the Newsan agent is initialised with a *source_trust* value of 0.5. The events are sent to a verification group consisting of n verifiers which verifies the content of the event and gives a { YES, NO, CANTSAY } reply. The modified equation for the calculation of the event trust is

$$E_j = \omega * S_{ij} + (1-\omega) * \left(1 - \alpha^{(n*10)/No.ofVerifiers}\right)$$

where the values of (ω, α, n) are $(0.6, 0.8\,, n)$ respectively and n is the positive responses from the verifier set.

An example of how an event is represented in RDF format and stored in the event model is shown below. The tuple *politician(001, N1, P3, Member, 06/02/2011, 16/11/2013*) is expanded to

(event has-eventid eventid:342)
(eventid generated-by source:s1)
(eventid described-as soc-agent:001)
(soc-agent:001 instance-of politician)
(soc-agent:001 has-name N1)
(soc-agent:005 instance-of political-party)
(soc-agent:005 has-name P3)
(soc-agent:001 has-state S1)
(S1 instance-of stative-sentence)
(S1 occurs_during TI1)
(TI1 start-date 06/02/2011)
(TI1 end-date 16/11/2013))
(soc-agent:001 member-of soc-agent:005)
(S1 subject soc-agent:001)
(S1 predicate member-of)
(S1 object soc-agent:005)
(eventid haseventTrust event_trust:0.5)
(event_trust equals 0.5)

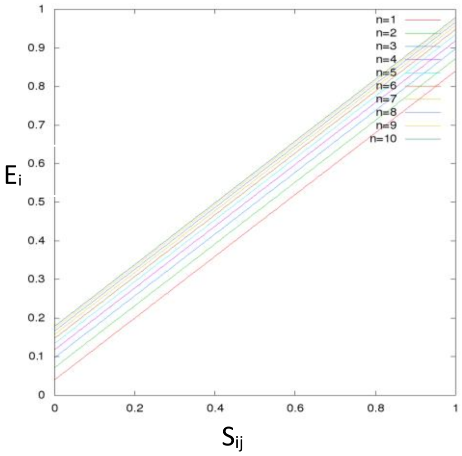

Figure 3 : $\omega = 0.6, \alpha = 0.8\ and\ 0 \leq n\ \leq 10$

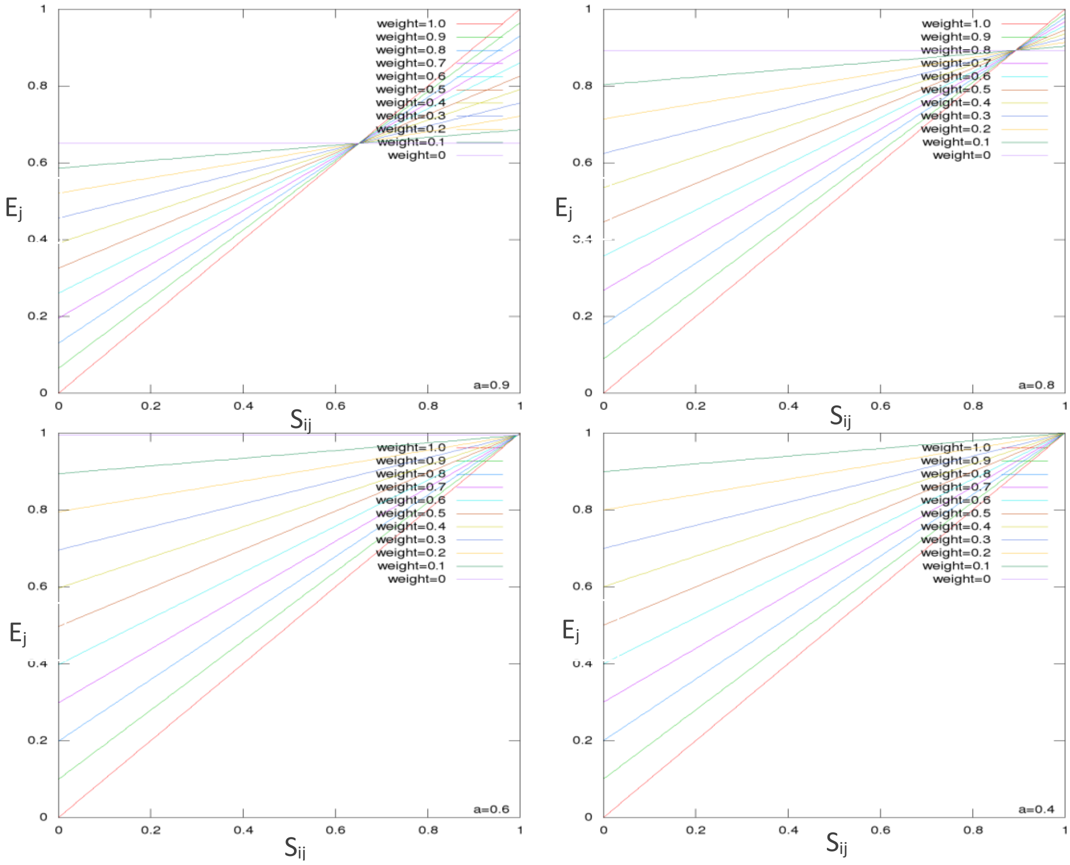

Figure 4 : (a) $\alpha = 0.9, n = 10$; (b) $\alpha = 0.8, n = 10$; (c) $\alpha = 0.6, n = 10$; (d) $\alpha = 0.4, n = 10$

4 CONCLUSION

Emerging information systems are increasingly dealing with real world happenings as captured by human being. They also capture and report experiential data such as audio records, photos, videos and from many other types of sensors. The necessity of storing all such information and also all its attributes and relationships with other objects and events are playing a vital role in real time query system. The main point of interest in all such events populated by various sources is how reliable is the event generated by such sources. In this dissertation work, we actually explored the different utilities required to quantify trustworthiness of the event and how we can manipulate it in a decentralized environment. It has been concluded that, one such approach is the E* graph based event model using RDF and ontologies incorporating a trust layer over this model. With the full power of modern day semantic relationships, the events can be queried and compared with ease.

5 BIBLIOGRAPHY

Bin Yu, M. P. (2004). *Developing Trust in Large-Scale Peer-to-Peer Systems.* IEEE First Symposium on Multi-Agent Security and Survivability.

C.Luckham, D. (2001). *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems.* Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Díaz, N. W. (1999). *Active database systems.* ACM Comput. Surv.

Duan, D. P. (1995). *An event-based spatiotemporal data model for temporal analysis of geographical data.* International Journal of Geographical Information Systems.

Huu Tran, M. H. (2005). *A Trust based Access Control Framework for P2P File-Sharing Systems.* Department of Computing, Macquarie University : Proceedings of the 38th Hawaii International Conference on System Sciences.

Gupta A., Jain R. (2011). *Managing Event Information - Modeling, Retrieval and Applications.* Morgan and Claypool.

Jain, U. W. (2007). *Toward a common event model for multimedia applications.* IEEEMultiMedia.