2010 International Conference on Complex, Intelligent and Software Intensive Systems

A Theoretical Multi-Tier Trust Framework for the Geospatial Domain

D. UMUHOZA[12], J.I. AGBINYA[1], A VAHED[2]

l Information and Communication Technology Group, University of Technology,
Sydney, NSW 2007, Australia;
({dumuhoza, agbinya}@eng.uts.edu.au)
2Meraka Institute, PO Box 395,
Pretoria 0001, South Africa; ({dumuhoza, avahed}@csir.co.za)

Abstract—**Sensory data also known as Geospatial data is collected in-situ or remotely by different types of sensors from different geographic locations, by different agencies and over a period of time. A vast amount of data is processed via web services. The quality of the techniques, models and algorithms applied along the processing chain in transforming sensory data critically influences the quality of the eventually derived information. In this paper we evaluate trustworthiness of the processing chain or workflow from data acquisition to knowledge discovery. We present work in progress of a theoretical multi-tier trust framework for processing chain from data acquisition to knowledge discovery in geospatial domain. Holistic trust will be computed trough a trust function that integrate take the existing trust models.**

Key-words: trust; workflow; Geospatial domain, processing chain.

I. INTRODUCTION AND PROBLEM DEFINITION

Trustworthy environmental sensing is of particular importance in the case of disaster management. For example, knowledge about the likelihood of a natural disaster such as an impending flood in a particular area comes from knowledge derived from meteorological, topographic and hydrologic data. Hydrographs at different geographic locations provide information about the rate and timing of stream flows and are indicative of other hydrological information such as river water levels, run-off and flow patterns. This type of data also known as Geospatial data is collected in-situ or remotely by different types of sensors from different geographic locations , by different agencies and over a period of time. The sharing of such resources allows for the possibility of re-using data and data analysis processes from various sites.

There are typically many complex steps and processes involved in transforming sensed data into discovered environmental knowledge. Acquired data such as remotely sensed imagery is transformed into useful information through well established processing chains that involve processes such as ortho-rectification, sensor calibration and atmospheric correction to compensate for sensing error and variation. Scientific techniques can also be applied to calibrate and validate acquired data before being transformed into final product presented to the user. One challenge is that a vast amount of data have to be processed since different users have different requirements, depending on the intended use of data [1]. Different technologies have to be applied in dealing with the huge amount of data and catering for different requirements and for different domains using distributed automated systems. One example of distributed processing is the application of grid computing for efficient sharing of resources.

The Sensor Web is a concept for accessing sensory resources and services published on the Internet and discovered as standards compliant service oriented architectures (SOA)[2] . Other examples of data resources include databases where environmental and geographic data and metadata are stored and retrieved as well as ontologies used to represent and reason about resources. Geospatial Web services are advantageous in this respect, as they can significantly reduce the volume of data and the computing resources required at the end-user side[3].

Integration of services and processes is realized via scientific workflows. These workflows are designed for methods to analyse data and solve problems that are specific to a particular scientific domain.

Clearly, the quality of the techniques, models and algorithms applied along the processing chain in transforming sensory data critically influences the quality of the eventually derived information.

The main focus of this paper is trustworthiness of the processing chain or workflow from data acquisition to knowledge discovery. Our particular interest is the knowledge that a user or a particular system has about the reliability, accuracy or correctness of the data or information as it is being transformed into knowledge. Figure 1 shows a summary of the processing chain which can also be represented as workflow from data acquisition to knowledge discovery.

In this paper we argue that trustworthiness of the information presented at the end user is derived from the reliability and correctness of each process involved. We also argue that reliability and correctness of processes are logically connected.

The first step in solving the problem is to track all processes involved in derivation of particular information or knowledge trough out the workflow. This tracking of origin of information or process is called provenance. Provenance

978-0-7695-3967-6/10 $26.00 © 2010 IEEE
DOI 10.1109/CISIS.2010.169

information sets the context of the information or knowledge to be analysed. It also provides means to validate experimental results by repeating the steps recorded in the provenance document. One, more or a combination of provenance methods suggested in the literature [4-6] can be used as a benchmark. We recognise that integration of these methods can also pose challenges and some have been outlined in [5]. We assume that a common way of describing provenance information for example by using SensorML can be a solution to this challenge as suggested in[7].

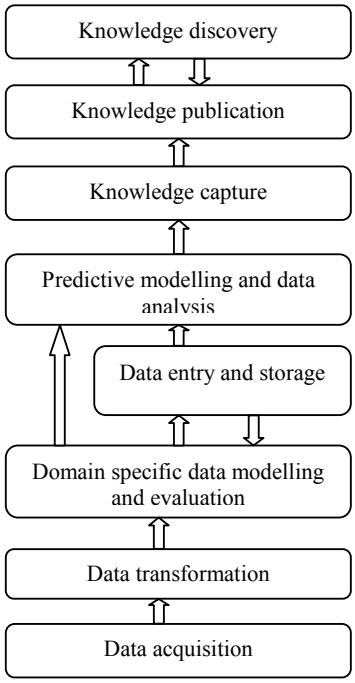

Figure 1: Main processes from data acquisition to knowledge discovery

We share the views in[5] that provenance information can also be used for quality control and reliability. However, we argue that the provenance framework must consider hereditary data and processes (which is sometimes referred to as data lineage) which is better illustrated through Figure 1. Provenance information can be used in the computation of trust thereby providing coherency of trust variables and processes as an improved approach for holistic trust evaluation.

We present a theoretical multi-tier trust framework using a bottom-up approach for processing chain from data acquisition to knowledge discovery in geospatial domain. Holistic trust will be computed trough a trust function that will take the existing trust models as input. The framework we propose can be applied in scientific workflow management systems. The framework provides the benefit of coherent decision making at different levels depending on the knowledge one/ system has or need about trust. We introduce the notion a combination of backward tracking of trust; where a decision is made based on the traversed *(present)* path and forward trust tracking where a decision is made based on the anticipated *(predicted)* path. The trust framework proposed is dynamic and adaptive.

The remainder of this paper is organized as follows. In section 2 we review and compare provenance techniques and trust models. Section 3 gives the overview of a multi-tier trust framework and relationships between the tiers of the framework. Section 4 describes a bottom-up approach for computing holistic trust. Section 5 gives discussions on the proposed trust framework. Section 6 provides conclusion and future work.

II. RELATED WORK

Trust definition and computation models have been developed for different purposes and goals. Trust for security in sensor networks have considered authentication methods and used credentials for evaluating trust [8, 9]. Trust for web services have used access policies as a way to determine the trusted service to access certain resources[10]. Trust based on quality of service has been used in web services and sensor networks [11]. Trust as is used in social networks has been used in agent systems and sensor networks based on recommendation systems[12, 13]. Trust definitions are therefore application and/or service dependent.

Trust models for Geospatial web services have been proposed with the goal of web service optimal selection or reputation and recommendation on Quality of Service of web service[10, 11, 13] or on quality of Geospatial data[14]. Trust policy for Semantic Web has also been proposed[15]. In this work we are interested in trust of workflows based on provenance information in Geospatial domain.

The scientific community uses different forms of provenance for experimental data and results. There have been a lot of activities on research about provenance knowledge in different contexts [4-6, 16]. Some researchers have described provenance in the context of database technologies[6], while others have described it in the context of experimental workflow[4, 16], and in the context of Geographic Information Science. All these contexts are related and the provenance their methods are comparable as the aim is to describe the origin and lineage of data and processes. The challenge is to integrate provenance information as that information is recorded for different uses, represented in different formats, stored differently with varying levels of detail depending on domains. Authors in [5] have proposed a provenance-integration framework for the Grid environment. They categorize provenance into internal provenance and external provenance. Existing internal provenance models and their corresponding repositories are wrapped in a web service. An index service framework is developed by collection provenance information with a workflow and provenance across workflows. The framework is used to map users' provenance request to corresponding provenance services.

Reasoning techniques can be applied to provenance information to deduce trust values but there has to be proper querying techniques for retrieving the provenance information.

There have been efforts to use provenance information in quantifying trust. Like in [17] provenance information was combined with workflow to evaluate trust on generated data. They suggest a provenance trust framework that takes three types of trust as input, namely, trust on process, trust on service and trust on data. These three types of trust are combined and an overall trust measure is determined. Provenance information used to assess trust is categorized in process provenance and actor provenance. Trust is evaluated through decision trees based on binary values using distributed probability and Bayesian analysis. A probable measure of trust is calculated based on different decision path values which are complete path, partial path, incomplete path and no-trust path.

In [18] the authors developed a fuzzy model for calculating workflow trust using provenance information as an improvement of the framework developed in [17]. The authors argue that it is possible to distinguish between an abstract workflow description and a concrete workflow description. They also argue that provenance information can be used to match the two types of workflows and deduce trust based on the results of the concrete workflow. They propose a fuzzy model to generate trust by considering trust as a subjective measure that can be of different degrees. A decision tree utilizes the provenance information obtained from a provenance store. An analysis path is defined in five steps which are process reliability, process relationship, data conflict, user requirements. The rules for evaluating trust are applied to all analysis nodes and produce real values in the [0,1] interval. Those values are the input to the fuzzy reasoning tool. Trust evaluation method in [17] and [18] provide an interesting way to use provenance information to compute trust. However these methods do not take into account workflow of workflow which is often the case for scientific workflows.

The authors of [19] argue that queries over lineage and uncertainty data must be evaluated subject to privacy and security constraints. They identified privacy, lineage, uncertainty and security as the most important aspects of information integration. The authors suggested a system termed PLUS (Synthesizing Privacy, Lineage, Uncertainty and Security). The PLUS system was designed considering a broad range of domains and applications. The common requirements identified are heterogeneity, lineage with workflows, bi-directional lineage traversal; which allows to reason backward and forward, variable granularity, incomplete disclosure, uncertainty, and poly instantiation. Based on the model in[19] we can assume that provenance information can be collected, stored and retrieved. Based on that assumption, we propose a holistic trust framework based on provenance information that is dynamic and adaptive. The framework takes the existing trust models as input to the holistic trust function. The framework will adapt to situations of missing or partial information and it will also adjust to particular conditions of users on their request.

III. MULTI-TIER TRUST FRAMEWORK

We introduce the multi-tier trust framework for sensory data, which consists of ten different related tiers of trust as depicted in Figures 2 and 3. Figure 2 presents trust at levels, while Figure 3 provides the basic relationships that exist between trusts (for example how trust A relates to trusts B, C and D). The framework is a holistic function of other trust functions that are related to each other in the framework. The holistic trust (T_h) can be defined as

$$T_h = T((A),(B),(C),\ldots,(J)) \quad (1)$$

Each trust represented by a label of alphabetical characters is a function that takes variable inputs and computes trust. The trust in each tier of the framework is calculated independently through the process path from data acquisition to knowledge discovery as depicted in Figure 2. Trust input variables in two or more tiers may be common. For example, the trust at points A and C may share delay as variable input.

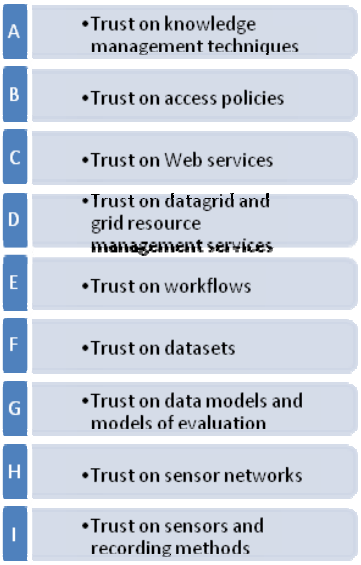

Figure 2: Multi-tier trust framework

3.1 TRUST RELATIONSHIPS

The components of the multi-tier trust framework represented in Figure 2 individually are input to the holistic trust function. However parameters of these trust functions may be related with each other and this could influence trust values as some of the trust function output serves as input to computation of the others trust function and vice versa. The relationships of how trust values influence each other can be depicted in a directed graph in Figure 3. In the Figure 3, the nodes represent trust function at given tiers and the directed arcs represent the relationship between the connected trust nodes.

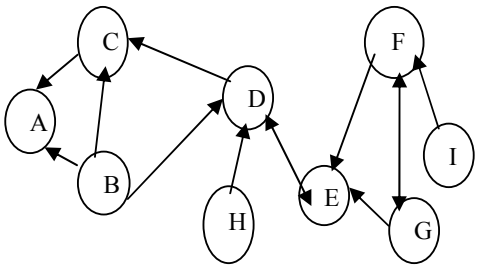

Figure 3: Trusts relationships graph

We consider each trust to be a function that takes a number of parameters as input. For example the trust function for trust A is given as follows:

$$A = f(a_1, \quad a_2, \quad \ldots, \quad a_N) \qquad (2)$$

where a_i $(1 \le i \le N)$ are the parameters for trust A. If two trust functions have similar or related parameters, the value of that parameter will be double counted in holistic trust function. Let us take an example of parameter delay. As delay of sensors and recording methods could influence or affect delay on sensor networks, the value of delay will be double counted when the two trust functions are integrated in order to compute holistic trust. In this case the problem of double counting can be avoided by determining the degree of similarities between parameters of different trust functions. We have intuitively represented the trusts relationships graph in Figure 3. But in practice, the graph will be build automatically as a result of reasoning algorithm that determines similarities between parameters.

We propose the following method to determine semantic similarities between parameters of trusts functions. We semantically describe trust functions and parameters in the form of ontology. Two main ontology classes are defined; class: TrustFunction and class: Parameters. Individuals are elements of each class. Each individual of TrustFunction class is linked to at least one individual of Parameter class by property "hasParameter". Individuals can be classes. In that case they are sub-classes of main class. For simplicity, in this paper we will limit our discussions to main classes and their individuals. For sake of readability, the remainder of this paper refers to individuals of a class as concept. Two concepts are related if they are linked to similar concepts. Suppose we have A and B as concepts of class TrustFunction, A is related to B if each one has as parameter an individual of class Parameter that measures delay. Generally we express similarity between concept A and concept B as follows:

(A) = all concepts of A , (B) = all concepts of B

S = All similar concepts between A and B.

$$(S) = \begin{cases} 1 \text{ if } (A) \cap (B) \neq \phi \\ 0 \text{ if } (A) \cap (B) = \phi \end{cases} \qquad (3)$$

$$Sim(A,B) = \frac{(S)}{(A)U(B)} \qquad (4)$$

Having established that two concepts are similar do not give us full information we need to deduct the level of similarity in our context of trust computation. We would like to establish how two concepts influence each other in order to compose forward or backward reasoning.

We find out how two similar concepts influence each other based on the directed graph G.

$G = (C, R)$ where C represents a set of concepts that we also call trust nodes. R represents a set of relationships between a pair (A,B) of concepts or nodes of concepts in C. Let us have two concepts A and B with their respective individuals $a_1, a_2, \ldots a_n$ and $b_1, b_2, \ldots b_n$ respectively. The overall strength of the relationship directed from between A to B will be determined by a relational similarity matrix $M_{G(AB)}$ where each element $a_i b_i$ of the matrix is a pair of nodes with a directed relationship from a_i concept to b_i concept. For all $1 \le i \le n$:

$$M_{G(AB)} = \begin{pmatrix} a_1b_1 & a_2b_1 & a_nb_1 \\ a_1b_2 & a_2b_2 & a_nb_2 \\ \ldots & \ldots & \ldots \\ a_1b_n & a_2b_n & a_nb_n \end{pmatrix} \qquad (5)$$

The value of each element of the matrix is computed with Equation 4. We classify the elements of the matrix in order to conclude on strength of the relationships of the graph G.

Because of uncertainties, imprecision and sometimes incomplete information that may occur while mapping trust functions into ontology, we choose to use fuzzy function to classify the relationship strength. We are aware that concepts could also be concepts of concepts. And in that case we can have a hierarchical taxonomy of graph and sub-graphs. In this paper we will limit our discussion on overall graph because of space.

In order to apply fuzzy membership function to determine the degree of similarity of the two concepts, we first transform the relational matrix $M_{G(AB)}$ into a fuzzy matrix $F_{G(AB)}$ using general fuzzy function[20]. Entries $a_i b_i$ of $M_{G(AB)}$ matrix are converted into fuzzy entries $c_i d_i$ using a fuzzifier

parameter α from interval [0,1], mean μb_i and the standard deviation σb_i of each b_i^{th} column as follows. For all $1 \le i \le n$:

$$\text{if } a_i b_i \le \mu b_i - \alpha * \sigma b_i \text{ then } c_i d_i = o$$

$$\text{else if } a_i b_i \in (\mu b_i - \alpha * \sigma b_i, \mu b_i + \alpha * \sigma b_i \quad (6)$$

$$\text{then } c_i d_i = \frac{a_i b_i - (\mu b_i - \alpha * \sigma b_i)}{(\mu b_i + \alpha * \sigma b_i) - (\mu b_i - \alpha * \sigma b_i)}$$

$$\text{else if } a_i b_i \ge \mu b_i + \alpha * \sigma b_i \text{ then } c_i d_i = 1$$

We have now defined three possible values of each element $c_i d_i$ in the matrix $F_{G(AB)}$. Therefore for each directed relationship between two concepts there are three possible classifications; strong, weak or non existent.

IV. TRUST COMPUTATION PROCESSES

We define a trust computation process framework. In this framework trust computation is a flow of processes that will run through four main components of the framework which are *consumer component, input component, and integration and evaluation component* as depicted in Figure 4.

Consumer: a consumer is either a user or a process (es). The consumer sets the parameters to be considered or requirements for the trust evaluation function. The consumer can use all available parameters or select a subset of few parameters to be used. These parameters are trust functions at different levels of the multi-tier trust framework in Figure 1.

Input: the selected parameters by the consumer are semantically mapped and interpreted in the same way using ontologies. A common language is used to describe each parameter and its values.

Integration: the selected trust functions are integrated. The integration will include the deduction of relationships of parameters in respective trust functions. Parameters and relations among them will be presented as a graph as shown in Figure3. The strength of the relationship is derived from the weight of the edges of the graph.

Holistic Trust computation: Holistic trust is a flow of process and takes different trusts functions and parameters as input. It is computed and evaluated in two phases where forward trust and backward trust are evaluated.- *forward trust* is computed based on the generalized initial consumer requirements. The reasoning engine fires rules based on the information contained in trust parameters memory. Based on the rules and values of parameters, it is deduced whether a process or data is trusted or not. If computed trust is going to be input to other trust functions, then further trust evaluation is required. – *Backward trust* is computed based on the anticipated path in the holistic trust workflow. The reasoning engine rule fires rules based on the specific requirements of a particular consumer. This process allows for refinement of matching consumer requirement. Hence as shown in Figure 4 backward trust evaluation takes forward trust value as input and checks if it will be acceptable and to what level by the consumer.

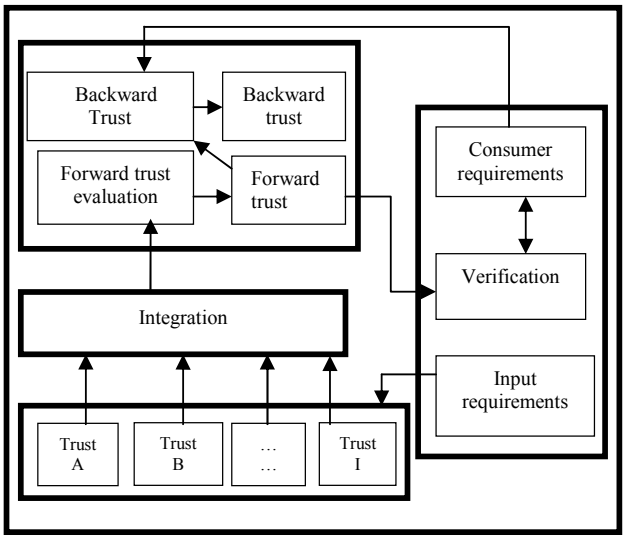

Figure 4: Forward-Backward Trust Computation Processes

Holistic trust function in Equation 1 will be implemented as complex workflow of trusts functions. Each trust function itself is a workflow of computations processes. Provenance of holistic trust workflow will be collected using modified provenance frameworks for scientific workflow systems. Provenance information will be parameters of trusts functions used to compute trust and rating methods used to measure trust.

Having trust workflow provenance information, we will be able to draw the directed graph in Figure 3. That graph will be the foundation of the Trust workflow model. Based on the graph we will formalize trust grammar using structural association rules and define syntax for the trust workflow using W3C standard like the Resource Description Framework (RDF). The trust grammar will allow common interpretation of trust values from heterogeneous sources.
The trust actor will be an aggregation of actors which will have a local director that can be implemented in an existing workflow management system like Kepler, taverna and others.

V. CONCLUSION AND FUTURE WORK

Holistic trust workflow is composed based on combination of backward and forward reasoning techniques. In the first phase forward reasoning will be done after integration of trusts functions. In the second phase backward reasoning will be done to evaluate the acceptability of the obtained trust in the first phase by the indented consumer of data or process of which trust is being measured. Combining the two reasoning techniques provides the flexibility and accuracy in decision making for the consumer to refine the trust requirements based on anticipated path at any stage during the trust workflow execution.

The trust workflow is easily adaptable to any geospatial application as the trusts inputs can be interpreted and aggregated as suggested in the trust process framework. This makes the trust framework suitable for a heterogenous and dynamic environment like the Sensor Web.

The framework can be adapted to other areas such as medical data and agricultural domains. Thus the system of trust computation has an inherent strength to re-adjust trust dynamically. In our continued research we will implement the proposed framework in a scientific workflow management system. We will test the framework with disaster management systems in case of flood. We will do performance analysis and draw conclusion and recommendation based on the results.

ACKNOWLEDGMENT

This work is supported by the Council for Scientific and Industrial Research (CSIR) and University of Technology, Sydney.

REFERENCES

[1] R. A. Schowengerdt, *Remote sensing: models and methods for image processing*: Elsevier Inc., 2007.

[2] G. Percivall, and C. Reed, "OGC® Sensor Web Enablement Standards," *On-line Magazine 'Sensors & Transducers' (S&T e-Digest),* vol. 71, no. 9, pp. 698-706, 2006.

[3] L. D. Peng Yue, Wenli Yang, Genong Yu, Peisheng Zhao, "Semantics-based automatic composition of geospatial Web serice chains," *Computers & Geosciences 33 (2007) 649–665*, 2006.

[4] S. B. Davidson, S. C. Boulakia, A. Eyal *et al.*, "Provenance in Scientific Workflow Systems," *IEEE Data Eng. Bull,* vol. 30, pp. 44-50, 2007.

[5] F. S. Jing Zhao, Carlo Torniai, Amol Bakshi, Viktor Prasanna, "A Provenance-Integration Framework for DistributedWorkflows in grid environment," *Workshop on Grid and Utility Computing,* vol. HiPC 2008 Workshops, December 17-20, 2008, 2008.

[6] W. C. Tan, "Provenance in Databases: Past, Current, and Future," *IEEE Data Engineering Bulletin, 30(4):3–12*, December 2007.

[7] M. B. J.J. Fredericks, T. Cook, J. Bosch, "Integrating Standards in Data QA/QC Into OpenGeospatial Consortium Sensor Observation Services," *Accessed online, September 2009*.

[8] B. Ma, "A Novel Stereoscopic Security Architecture with Trust Management for Wireless Sensor Networks," *Ist International conference on Communication Software and Networks*, 2009.

[9] E. R. d. Mello, M. S. Wangham, J. d. S. Fraga *et al.*, *A Model for Authentication Credentials Translation in Service Oriented Architecture* p.^pp. 68-86, Heidelberg: Springer Berlin 2009.

[10] Y. Gil, and D. Artz, "Towards content trust of web resources," *Proceedings of the 15th international conference on World Wide Web*, pp. 565 - 574 2006

[11] P. Zedan, and J. Baik, "*QoS Broker-Based Trust Model for Effective Web Service Selection*," *IASTED SEA,* vol. Cambridge, Massachusetts, USA, 2007.

[12] I. R. Josang A. , and Boyd C., "A survey of trust and reputation systems for online service provision," *Decision Support Systems,* vol. 43, pp. 618 - 644,, 2005.

[13] Y. Wang, and J. Vassileva, "*A review on trust and reputation for web service selection*," *1st Int. Workshop on Trust and Reputation Management in Massively Distributed Computing Systems*, 2007.

[14] G. Subbiah, A. Alam, L. Khan *et al.*, "*Geospatial Data Qualities as Web Services Performance Metrics*," *15th International Symposium on Advances in Geographic Information Systems*, 2007.

[15] S.Galizia, A. Gugliotta, and J. Domingue, "A Trust Based Methodology for Web Service Selection," *International Conference on Semantic Computing,*, 2007.

[16] O. B. Ilkay Altintas, Efrat Jaeger-Frank, "Provenance collection support in the kepler scientific workflow system," *Provenance and Annotation of Data*, pp. 118--132, 2006.

[17] S. Rajbhandari, I. Wootten, A. S. Ali *et al.*, "Evaluating Provenance-based Trust for Scientific Workflows," *Sixth IEEE International Symposium on Cluster Computing and the Grid.* , 2006.

[18] S. Rajbhandari, O. F. Rana, and I. Wootten, "A fuzzy model for calculating workflow trust using provenance data " *Proceedings of the 15th ACM Mardi Gras conference* 2008.

[19] B. Blaustein, L. Seligman, M. Morse *et al.*, "PLUS: Synthesizing privacy, lineage, uncertainty and security," *IEEE 24th International Conference on Data Engineering Workshop.*, 2008.

[20] *First Course on Fuzzy Theory and Applications*: Springer Berlin / Heidelberg, 2006.