# Data centric trust evaluation and prediction framework for IOT

JAYASINGHE, Upal, OTEBOLAKU, Abayomi <http://orcid.org/0000-0002-4320-9061>, UM, Tai-Won and LEE, Gyu Myoung



**Published version**

JAYASINGHE, Upal, OTEBOLAKU, Abayomi, UM, Tai-Won and LEE, Gyu Myoung (2018). Data centric trust evaluation and prediction framework for IOT. In: 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K). IEEE, 1-7.

# DATA CENTRIC TRUST EVALUATION AND PREDICTION FRAMEWORK FOR IOT

*Upul Jayasinghe[1], Abayomi Otebolaku[1], Tai-Won Um[2], Gyu Myoung Lee[1]*

[1]Department of Computer Science, Liverpool John Moores University, Liverpool, L3 3AF, UK.
[2]Department of Information and Communication Engineering, Chosun University, Gwangju, Korea.
u.u.jayasinghe@2015.ljmu.ac.uk, a.m.otebolaku@ljmu.ac.uk, twum@chosun.ac.kr, g.m.lee@ljmu.ac.uk

## ABSTRACT

*Application of trust principals in internet of things (IoT) has allowed to provide more trustworthy services among the corresponding stakeholders. The most common method of assessing trust in IoT applications is to estimate trust level of the end entities (entity-centric) relative to the trustor. In these systems, trust level of the data is assumed to be the same as the trust level of the data source. However, most of the IoT based systems are data centric and operate in dynamic environments, which need immediate actions without waiting for a trust report from end entities. We address this challenge by extending our previous proposals on trust establishment for entities based on their reputation, experience and knowledge, to trust estimation of data items [1-3]. First, we present a hybrid trust framework for evaluating both data trust and entity trust, which will be enhanced as a standardization for future data driven society. The modules including data trust metric extraction, data trust aggregation, evaluation and prediction are elaborated inside the proposed framework. Finally, a possible design model is described to implement the proposed ideas.*

***Keywords***— Data Trust, Knowledge, Reputation, Experience, Collaborative Filtering, Ensemble Learning.

## 1. INTRODUCTION

With the exponential growth of applications of internet of things (IoT) including social networks and e-commerce systems, users always surf in the universe of data, in which users often do not know about who they are interacting with and receiving data from. In such situations, the concept of trust plays an important role in managing these interactions and developing a trustworthy environment for all providers, users and the communities. However, generating trust relationships among users is extremely hard due to diversified nature of the users and how each entity understands trust. In traditional forms of trust management systems, trust is computed based on the relationship among end entities and behaviors in certain transactions as explained in [1], [2], [4-6]. Moreover, these systems use certain set of metrics like honesty, cooperativeness, community interest, reputation, certificate validity, length/frequency of the transaction and etc., to evaluate the trustworthiness of end entities and then to find trust relationship among the trustors and the trustees.

However, trust on end entities is not always prominent but the data receiving in form of various types. As an example, reliable, up to date and location sensitive information about weather, traffic, safety warnings and transport information from a smart city application are more important than the facts about entities who are actually generating them. The other common misinterpretation is that the assumption of having entity trust would guarantee data trust which is in fact indubitably different in various aspects like validity of data, timeliness and other properties unique to data which are often ignored in calculating trust for end entities. Further, information is the governing factor for any IoT systems and is generated from the data by combining it (data) with the context. Hence, if there is a data quality (DQ) problem, it would eventually lead to information quality (IQ) problem [22]. In other words, once the right data item is delivered to a desired entity at the precise time in a clear, useable and meaningful manner, IQ is guaranteed.

Therefore, it is important to address the challenge of establishing a data centric trust while preserving the traditional form of trust computation. To this end, firstly we define a set of dynamic factors, which essentially describe the DQ attributes and also metrics which define the knowledge, experience and reputation as in our previous work [2] and [1] to get the best of traditional means of trust computation. Then, we combine these attributes built on REK (Reputation, Experience and Knowledge) model described in [4] and [7]. After that, a technique which assesses the data centric trust for every user who is new to the system and who needs to access the data streams, is investigated based on the concepts of recommendation systems (RS). Here, we apply the RS due to its ability to generate approximate trust value for unknown records based on the available trustor–trustee relationships. Finally, we discuss a realistic design model of the proposed items.

From global standardization perspective, ITU-T Study Group (SG) 13 established the correspondence group on trust (CG-Trust) for preliminary work on trust standardization [8]. The CG-Trust developed a technical report containing definition, use cases, functional classification as well as challenges, technical issues related to trust including overall strategies of standardization for trust provisioning. As the lead group of trusted networking infrastructure, ITU-T SG13 successfully completed to publish the recommendation Y.3052 on trust in March 2017 [9]. Recently Question 16/13 "*Knowledge-centric trustworthy networking and services*" has focused on basic issues and key features on trust. Q16/13 is now mainly focusing on the development of core technical solutions for trust provisioning from ICT infrastructures and services. Q16/13 also plans to consider technology

deployment as well as new services and business aspects on trust-based networks and eco-platforms. Our proposals provide a strong suggestion to improve the current standardization activities on trust in ITU-T SG13 towards a hybrid model based on the concepts of entity trust and data trust. Among them, a trust relationship model described in this work elaborates some important factors when it comes to trust based decision making that is a vital part of standardization process. On the other hand, trust evaluation via ensemble methods, which is by combining numerical, machine learning and recommendation algorithms, provides robust perception about trust compared to traditional one dimensional trust calculation techniques [5-7]. Additionally, we propose and encourage to use publish-subscribe architecture in the process of data management due to its distributed and autonomous nature.

The rest of this paper is structured as follows. Section 2 provides a comprehensive overview of the related research that has been conducted in relation to trust assessment and prediction. Section 3 confers a generic trust assessment architecture. Based on the generic model described, Section 4 proposes a numerical model for preliminary data trust computation and trusted data prediction. A possible implementation scenario is explained in Section 5 and Section 6 concludes the paper and outlines our future work.

## 2. RELATED WORK

Marsh proposed "Formalizing trust as a computational concept" [10] and argued that trust is the degree of uncertainty and optimism regarding an outcome. He further explains a trust model based on three trust metrics, direct trust, trust based on experience and the situational trust. Even though the direct trust measurements are the most reliable way of assessing trust, when it comes to applications like social networking, indirect measurements are more prominent due to collaborative behavior of the users. In this sense, [11], [12] and [2] discuss trust assessment models based on indirect trust metrics like reputation and recommendation. Further, there are situations when both direct and indirect trust information are not available. In such situation, "stereo-trust" [13] will be appropriate to generate first guess of trustworthiness even before the direct interactions occur.

Moreover, social interactions among entities disclose the valuable information of trust in analogy to the sociology concept of human interactions based on trust relationships. [14-17] discuss such models based on the social trust metrics like community of interest, friendship, followers, and frequency/duration of an interaction. After trust metrics are calculated individually, it is a must to combine them to have an overall idea about the final trust value. [18] and [19] investigate such a model based on the adaptive weightages. However, assessment of a proper weightage is a complex task due to the fact that trust is a varying quantity which depends on many factors like expectations of a trustor, time, context, etc. Thus, more intelligent schemes are required, preferably with well-known machine learning techniques like regression, supervised and unsupervised learning as

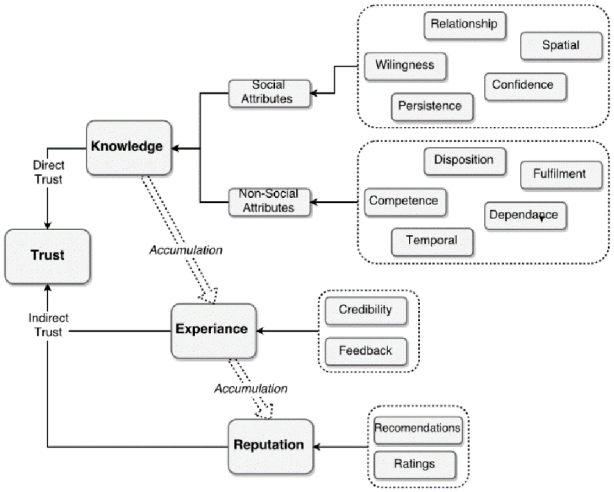


Fig. 1. A Generic Trust Model.

discussed in [5], [20] and [21]. Presently, data is the key governing factor with respect to service provisioning and decision making process in IoT. Hence, the assurance of DQ and IQ are utmost important for trustworthy interactions. In this regard, authors in [22], [23] and [24] discuss various techniques and metrics that can be considered for DQ and IQ measurement. The framework proposed by Askham *et al.* [25] is one of the most prominent and widely accepted model for DQ assessment due to its generic nature. Hence, we adopt most of the concepts from this work in order to develop our framework. Moreover, authors in [26] and [27] argue a data centric trust model for vehicular networks based on several techniques like Bayesian inference and Dempster-Shafer theory.

In contrast to traditional means, there are several work on trust prediction based on collective methods where numerically assessed trust metrics are analyzed through an intelligent algorithms like supervised and unsupervised learning. In this regard, a model to improve trust prediction accuracy by combining user similarity rating and the traditional trust is proposed in [28] and [29]. Furthermore, Xiang *et al.* proposes a model based on unsupervised learning algorithm to estimate relationship strength from interaction activities like tagging, communication and interference [30].

## 3. TRUST IN IOT

Among the various definitions of trust, we identify trust as a qualitative or quantitative property of a trustee measured by a trustor for a given task in a specific context and in a specific time period [1]. Furthermore, we distinguish properties of trustworthiness into three categories: Reputation, Experience, and Knowledge as we proposed in [3], [1] and [4] and formulate a trust assessment model as shown in Fig. 1. The Knowledge trust metric (TM) incorporates the first party or direct information, provided by a trustee to evaluate its trustworthiness and estimated by some trust attributes (TAs) depending on the services and entities. As examples, relationship attributes (Co-location, Co-work and parental), cooperativeness, spatial attributes (social centrality, community of interest) and temporal attributes (frequency

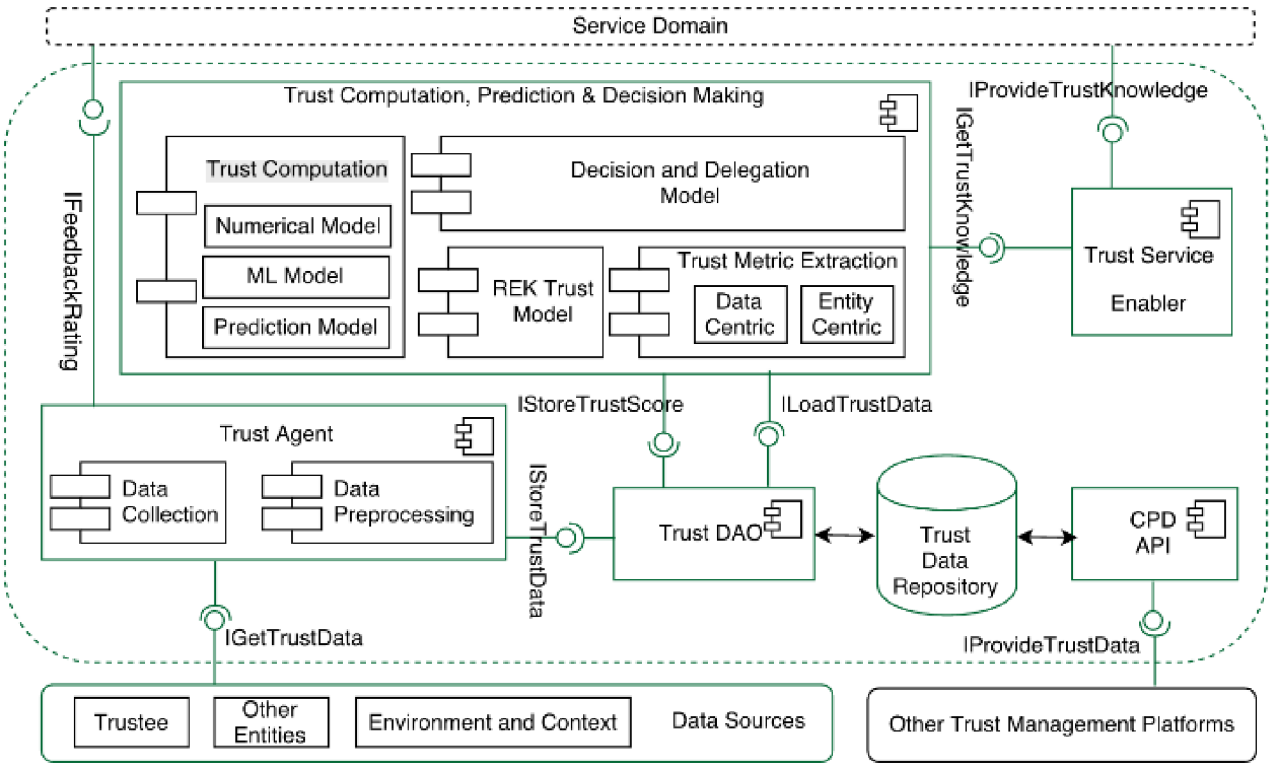


Fig. 2. Data Centric Trust Evaluation and Prediction Framework.

and duration of interactions) can be recognized as some of prospective TAs for knowledge TM.

Moreover, the main purposes of trust assessment are to facilitate more intelligent decision making and task delegation. In this regard, we further elaborate two more metrics, which comes under knowledge TM as non-social TMs and social TMs. In non-social trust, the idea is to find whether the trustor can rely on a physical or cyber entities and social trust determines whether a trustor can depend on other social entities [14]. We define four parameters; Competence, Disposition, Dependence and Fulfilment, which define the non-social trust as well as three parameters; Willingness, Persistence and Confidence which define the social trust when it comes to delegation and decision making as opposed to believes discussed in [28]. With respect to the REK model, these additional metrics define the knowledge TM particularly in the decision making process. Let us consider a specific trustor *A* and a trustee *B* with respect to a particular goal *g* in the decision making process. Based on this setup the definitions of the aforementioned attributes are;

- Competence Trust: B is beneficial and capable of realizing *g*
- Disposition Trust: B actually performs the task
- Dependence Trust: Achievement of goal *g* relies upon B
- Fulfilment Trust: B's contribution is necessary to achieve the task
- Willingness Trust: B shows no resistance over accomplishing the goal *g*
- Persistence Trust: Consistency over time, conquering the task
- Confidence Trust: Confident about B himself towards realizing *g* [31]

In the meantime, reputation and experience TMs falls under indirect observations as information to calculate these metrics comes only after a particular interaction or from third party sources. The process of indirect trust measurement is essentially an interactive process as shown in Fig. 2. For instance, attributes such as credibility and feedback which represent experience metric can be calculated only with the accumulated knowledge metrics. Similarly ratings and recommendations can only be generated after the accumulation of experience over a community.

## 4. DATA TRUST FRAMEWORK

To the present day, evaluation of trust in data is assumed to be identical to trust estimation of end entities. However, this is not entirely true and in fact most IoT systems rely highly on several data streams and these systems often care about the integrity and quality of who is generating them. As an example, obtaining accurate information about certain accident situation from less trustworthy entities like taxi drivers and passengers are more important than waiting for a report from a police officer, who is more trustworthy to a taxi driver, in order to get quick attention from medical authorities and other relevant parties. Another example is where the interactions happened for short duration without any prior relationship with the trustee. In such situations, it can be a disadvantage to calculate trust between entities due to time criticalness of the application.

To address these challenges, we propose a Data Centric Trust Evaluation and Prediction Framework as shown in the Fig. 2, which is capable of analyzing both data centric as well as

entity centric trust separately or in a collective manner. The platform consists of several important modules such as Trust Computation, Prediction and Decision Making (TCPD), Trust Agent (TAg), Trust Data Access Object (TrustDAO), Data Repository, Trust Computation and Decision making module, Trust Service Enabler and API. Once the TCPD identify a requirement of data, it ask the TAg via Trust DAO to collect necessary information and preprocessed them for trust evaluation. Then, these preprocessed data is stored in the data repository to be used by other modules including external platforms through TCPD API.

Afterwards, trust metric extraction module estimates the necessary trust attributes based on the requirement. These attributes can either be categorized as data centric attributes as explained in Section 4.1 or traditional entity centric attributes as described in Section 3. Next, all the attributes are combined based on the REK model with the assistance of trust computation module, which is capable of performing the calculation based on either numerical methods or artificial intelligence approach as described in sections 4.2 and 4.3 respectively. Finally, decision making and delegation module uses the predicted trust values in order to complete the decision process perhaps with the support of service enabler who is actually perform the judgement made by the decision module. In the following sections, we explain the data centric trust attribute estimation, data trust computation and data trust prediction in detail.

### 4.1. Data Trust Attributes

Alongside with the REK model, we first consider a separate set of trust attributes which essentially define the properties of data. Many research work on DQ shows that the six parameters (e.g., completeness, uniqueness, timeliness, validity, accuracy and consistency) provide prominent insight for assessing the DQ matters as in [22], [25], [32]. With respect to trust notion, we can consider these properties as trustworthiness attributes. Further, we consider two additional attributes, "success" and "cost", which characterize experience and reputation data trust metric (DTM) calculation, in additional to aforementioned attributes stated in Section 3. We consider these eight data trust attributes (DTA) as the core dimensions in finding the trust between a data item and the trustor. Thus, we model these properties as below:

- Success ( $T_B^{su}$) : the probability that B will successfully execute the task
- Cost ($T_B^{ct}$) : the probability that the cost of executing the task by B is not more than expected
- Completeness ($T_B^{cm}$): the probability of complete data records over total data records
- Uniqueness ( $T_B^{uq}$ ): the probability of expected records over total records noted
- Timeliness ( $T_B^{tm}$ ): the difference between last update to the current one
- Validity ($T_B^{vl}$): the validity of data type, syntax and range
- Accuracy ($T_B^{ac}$): the probability of accurate data records over total data records

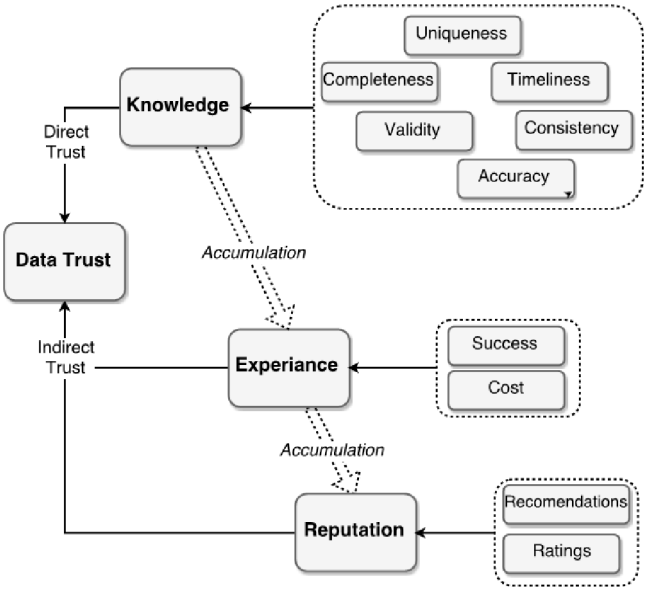


Fig. 3. A Data Trust Model.

- Consistency ( $T_B^{cn}$ ): the probability of valid, accurate and unique records over total data records

### 4.2. Data Trust Computational Model

In this section, we extend our entity centric model in Fig. 1 to comply with the data centric trust as shown in Fig. 3 and explain how each DTA is combined to generate data centric trust. For that we identify completeness, uniqueness, timeliness, validity, accuracy and consistency as DTAs which represent knowledge TM as it conveys trustworthiness information about the trustee. On the other hand "success" DTA and "cost" DTA represent the experience DTM of the trustor after each task. Finally reputation DTM can be considered by aggregating opinions of other trustees if there are any. Based on this, basic data trust assessment towards B by A ($T_{AB}^{x}$) over $x$ DTM can be numerically modeled as below:

- Knowledge DTM ($T_{AB}^{K}$)

$$T_{AB}^{K} = \alpha T_B^{cm} + \beta T_B^{uq} + \gamma T_B^{tm} + \delta T_B^{vl} + \varepsilon T_B^{ac} + \epsilon T_B^{cn} \quad (1)$$

  where α,β,γ,δ,∈, and ε are weighting factors such that α+β+γ+δ+∈+ε=1. However, calculating these weighting factors are computationally costly and not practical due to infinite number of possibilities. Hence, we suggest to apply machine learning (ML) techniques to combine all TAs, which we have discussed in our previous work [7].

- Experience DTM ($T_{AB}^{E}$)

$$T_{AB}^{E} = \sigma T_B^{su} + \varphi \frac{1}{T_B^{ct}} \quad (2)$$

  where σ and ϕ are weighting factors such that σ+ϕ =1 and $T_B^{ct} > 0$. The ML method discussed in [7] is preferable for TA combination in this case as well.

– Reputation DTM ($T_{AB}^{R}$)

$$T_{AB}^{R} = T_{1B}^{R} + T_{2B}^{R} + \cdots + T_{nB}^{R} \quad (3)$$

where $T_{nB}^{R}$ represents the reputation towards data source B by its previous users *n*. A mechanism that computes reputation based on PageRank algorithm is presented in our previous research [2].

After releasing the main DTMs, the next objective is to combine them in order to produce a final data trust value ($T_{AB}^{d}$) for each data source based on DTAs as below:

$$T_{AB}^{d} = \rho T_{AB}^{K} + \tau T_{AB}^{E} + \omega T_{AB}^{R} \quad (4)$$

where ρ,τ, and ω are weighting factors based on the trustors preference on each TM. In here, we suggest two mechanisms to combine each TM either based on the ML approach we followed in [7] or applying the rule based reasoning mechanism explained in [4].

### 4.3. Data Trust Prediction

Once the trust values based on DTA are collected, next step is to find the trust relationship among data sources and the trustors who do not have prior encounters. For that, we use the concepts of well-known collaborative filtering (CF) technique to predict the unknown trust values between the user and specific data source with respect to six different data centric features (e.g., completeness, uniqueness, timeliness, validity, accuracy and consistency). As now the predication is solely based on properties of data, it is unnecessary to rely on trustworthiness of the data source as in traditional methods anymore. Among various methods of recommendation techniques, we particularly choose a variant of a multifaceted CF model for our application due to its unique properties that match with our data trust model like stressing the concept of social contribution where everyone's contribution matters, capacity to capture weak signals in the overall data, ability to detect strong relationships between close items and competence to avoid overfitting [33].

First, we define the inputs to our algorithm as number of trustors or users ($n_u$), number of Trustees or DSs ($n_m$) and six features as shown in Table 1. Users who already have trust relationship with DSs are noted with "◬" symbol which actually represents some trust value between [0,1], calculated using equation (4) and the blank spaces denote the missing information, which is to be predicted. Formally, if user *j* and item *i* already have trust relationship, then r(*i,j*)=1 and r(*i,j*)=0, otherwise. Moreover, the data trust value given by user *j* to DS *i* is denoted by $y^{(i,j)}$. The symbol "◈" represents the values of each six features in between 0 and 1.

The next step of our algorithm is to find a parameter that describes the profile of users involved in a certain situation. For now let's assume this parameter is denoted by $\boldsymbol{\theta}^{(j)}$ for a particular user *j* and feature vector for DS *i* is denoted by $\mathbf{T}^{(i)}$. Then the predicted data trust value $T^{dp}_{ij}$ between the trustor and the data can be calculated as in equation (5). The symbol $(.)^T$ represent the transpose of the vector.

$$T_{ij}^{dp} = \left(\boldsymbol{\theta}^{(j)}\right)^T \left(\boldsymbol{T}^{(i)}\right) \quad (5)$$

Table 1. The users × items × features input matrix of the CF algorithm.

| | Trustors (Users) | | | | Features | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Trustees (DS) | $u_1$ | $u_2$ | … | $u_{nu}$ | $T^{cm}$ | $T^{uq}$ | $T^{tm}$ | $T^{vl}$ | $T^{ac}$ | $T^{cn}$ |
| $i_1$ | ◬ | | ◬ | | ◈ | ◈ | ◈ | ◈ | ◈ | ◈ |
| ⋮ | | ◬ | | ◬ | ◈ | ◈ | ◈ | ◈ | ◈ | ◈ |
| $j_{nm}$ | | ◬ | ◬ | | ◈ | ◈ | ◈ | ◈ | ◈ | ◈ |

The basic but essential requirement of the predicted trust value is that it must provide closest possible prediction for each trust value that is already calculated by each user. With this assumption, we can use mean square error (MSE) method to find the distance between actual trust values and predicted one. The parameter $\boldsymbol{\theta}^{(j)}$ which gives minimum error would be our best predicted trust value. This idea is formulated as below for trustor *j*:

$$\min_{\boldsymbol{\theta}^{(j)}} \frac{1}{2}\sum_{i:r(i,j)=1}\left(\left(\boldsymbol{\theta}^{(j)}\right)^T\left(\boldsymbol{T}^{(i)}\right) - y^{(i,j)}\right)^2 + \frac{\lambda}{2}\sum_{k=1}^{6}\left(\boldsymbol{\theta}_k^{(j)}\right)^2 \quad (6)$$

In the first part of the equation, the mean error is calculated over all the records where the trust value is already available through preliminary calculation. The second part of the equation is used to regularize the minimization process and there-by avoiding the overfitting issues. The *k* denotes the number of features. Similar manner, we can find the best parameter for each trustor as below:

$$\min_{\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},\ldots,\boldsymbol{\theta}^{(n_u)}} J(\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},\ldots,\boldsymbol{\theta}^{(n_u)}) \quad (7)$$

where J(.) denotes the cost function as described in equation (6). In order to minimize the cost function, we simply adapt the gradient decent method and solve for best parameter $\boldsymbol{\theta}^{(j)}_k$ as below [34]:

$$\theta_k^{(j)} = \begin{cases} \theta_k^{(j)} - \alpha\sum_{i:r(i,j)=1}\left(\left(\boldsymbol{\theta}^{(j)}\right)^T\left(\boldsymbol{T}^{(i)}\right) - y^{(i,j)}\right)^2 T_k^{(i)}, k = 0 \\ \theta_k^{(j)} - \alpha(\sum_{i:r(i,j)=1}\left(\left(\boldsymbol{\theta}^{(j)}\right)^T\left(\boldsymbol{T}^{(i)}\right) - y^{(i,j)}\right)^2 T_k^{(i)} + \lambda\theta_k^{(j)}), k \neq 0 \end{cases} \quad (8)$$

Once the parameter $\boldsymbol{\theta}^{(j)}$ is estimated through equation (7) and (8), predicted trust value between user *j* and item *i* will be given by the equation (5). Please note that this process is an iterative process and that more users who have experience with similar DSs would make the system more accurate and trustworthy.

## 5. IMPLEMENTATION MODEL

In this section, we propose a possible implementation scenario of our findings based on air pollution crowd sensing use case, aimed at collecting and monitoring pollution data. The air pollution sensing requires active citizen participation by carrying wearable sensors as they traverse the city based on opportunistic crowd sensing application [35]. However, monitoring such air pollution via crowd sensing requires that the data being provided are trustworthy and can be relied upon by city authority or government to make an immediate decision. The air pollution crowd sensing application will take advantage of citizen's smartphones and smart city's air

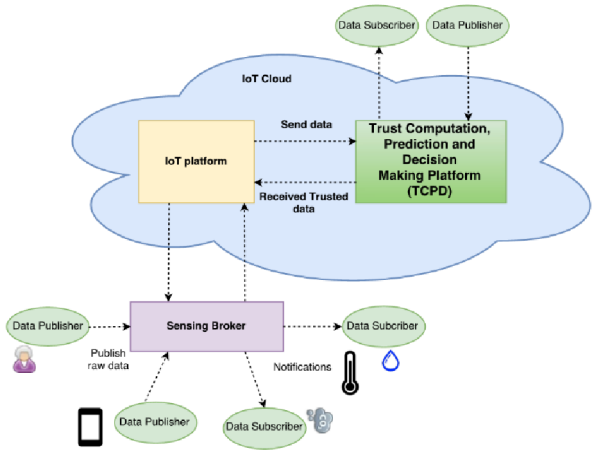


Fig. 4. High-level Implementation Architecture of the proposed System.

pollution/environment sensors. The data collected from the air pollution sensors are delivered to the IoT Cloud, hosting the TCPD proposed in this paper. Thus, a mobile app for trusted air quality data monitoring can be developed on top of this framework integrating data collected from low-cost environment sensors for temperature, humidity, CO, CO2 NO2, SO2, as well as compounds including benzene and lead (VOCs), etc. The sensors' readings will be transmitted via either an Android or IOS app to the proposed system for assessing and predicting the trust of the data before it is sent to the IoT Cloud. Such data can then be visualized along with its trust level by interested individuals, government, city administrators etc. via a web application.

For the above use case to profit from the proposed solution, we have proposed a distributed publish-subscribe architecture such as CoreDX distributed publish subscribe middleware [36] whereby an interested parties can subscribe via a broker to environmental data of interest in specific location of their choice as illustrated in Fig. 4, the implementation architecture. TCPD section of the figure implements appropriate components of the framework as shown in Fig. 2, for providing trusted data to the interested parties. This is a typical publish subscribe system whereby publishers publish the sensor data to the broker and subscribers receive notifications matching their subscriptions from the broker. As illustrated in the Fig. 2, the TCPD can communicate with the IoT platform via an edge server that implements the IGetTrustData and IProvideTrustData interfaces. Also, the TCPD can receive data from the IoT platform for predicting the trust of such received data.

Finally, Fig. 5 illustrates an example of a scaled down sequences of interactions between some important stakeholders of an implementation instance of the system. Anytime a new environment sensor is available, it registers its presence with the sensing broker, which in turn informs the framework of the new available sensor. The new sensor can then publish its data to the broker. The broker notifies the TCPD to predict the trust of the received data. Similarly, whenever a new subscriber joins the system, its subscription is submitted to the broker via the TCPD system. If a subscription matching at least one of the subscriptions of the

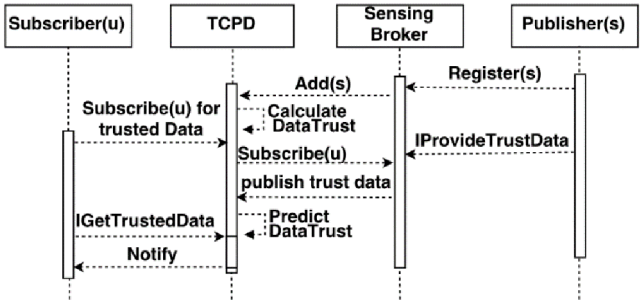


Fig. 5. Scaled down Sequence Diagram Showing relationship among the Publisher (Sensor), Subscriber (u), Sensing Broker and the proposed Framework.

new subscriber is available, the broker notifies the TCPD system to deliver the data to the subscriber along with the trust level of the data.

## 6. CONCLUSION

In this work, we argue that the traditional means of trust computation for entities does not necessary guarantee the trustworthiness of data that they generate. Hence, we propose a hybrid trust computational platform which is capable of assessing both data centric trust as well as traditional entity based trust. Further, we provide a model to compute individual DTA and the main DTM by combining numerical models with learning algorithms. Afterward, a data trust prediction scheme based on collaborative filtering is proposed to find the data trust between trustors and data sources who do not have prior encounters that avoids using data from malicious actors. Finally, a possible implementation scenario is discussed based on a crowd sensing use case. Similarly, our algorithm would be beneficial to filter out malicious data and data sources to maintain integrity and quality of the outcomes that any crowd sensing application produces.

For future work, we would like to incorporate content and contextual information for data trust prediction and propose a more accurate prediction model based on artificial intelligence concepts. Although ITU-T has started a new work on trust index which is a comprehensive accumulation of trust indicators to evaluate and quantify trust of entities, until now, standards on trusted data are still very limited and current standards on entity or network based trust must be expanded for taking into consideration the data trust matters as explained in this work.

## ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT). [2015-0-00533, Development of TII(Trusted Information Infrastructure) S/W Framework for Realizing Trustworthy IoT Eco-system].

## REFERENCES

[1] U. Jayasinghe, H. W. Lee, and G. M. Lee, "A Computational Model to Evaluate Honesty in Social Internet of Things," in *32nd ACM SIGAPP Symposium On Applied Computing*, Marrakesh, Morocco., 2017.

[2] U. Jayasinghe, N. B. Truong, G. M. Lee, and T.-W. Um, "RpR: A Trust Computation Model for Social Internet of Things," in *2016 Intl IEEE Conference on Smart World Congress*, Toulouse, France, 2016.
[3] N. B. Truong, G. M. Lee, and T.-W. Um, "A Reputation and Knowledge Based Trust Service Platform for Trustworthy Social Internet of Things," in *Innovations in Clouds, Internet and Networks (ICIN)*, Paris, France., 2016.
[4] N. B. Truong, H. Lee, B. Askwith, and G. M. Lee, "Toward a Trust Evaluation Mechanism in the Social Internet of Things," *Sensors,* vol. 17, no. 6, pp. 1346, 2017.
[5] Y. Wang, Y.-C. Lu, I.-R. Chen, J.-H. Cho, A. Swami, and C.-T. Lu, "LogitTrust: A Logit Regression-based Trust Model for Mobile Ad Hoc Networks," in *Proceedings of the 6th ASE International Conference on Privacy, Security, Risk and Trust* Cambridge, MA, 2014, pp. 1-10.
[6] I. R. Chen, F. Bao, and J. Guo, "Trust-based Service Management for Social Internet of Things Systems," *IEEE Transactions on Dependable and Secure Computing,* vol. PP, no. 99, pp. 1-1, 2015.
[7] U. Jayasinghe, G. M. Lee, T.-W. Um, and Q. Shi, "Machine-Learning-based Trust Computational Model for Social IoT Services," *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, Under Review, 2017.
[8] ITU-T SG13. "Future networks, with focus on IMT-2020, cloud computing and trusted network infrastructures," https://www.itu.int/en/ITU-T/studygroups/2017-2020/13/Pages/default.aspx.
[9] ITU-T Y.3052, "Overview of trust provisioning for information and communication technology infrastructures and services," March 2017, https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13252.
[10] S. P. Marsh, "Formalising trust as a computational concept," Ph.D. dissertation, Dept. Computing Science and Mathematics, University of Stirling, Stirling, Scotland, UK., 1994.
[11] A. Josang, and R. Ismail, "The beta reputation system." in Proceedings of the 15th bled electronic commerce conference, vol. 5, pp. 2502-2511, 2002.
[12] L. Xiong, and L. Liu, "Peertrust: Supporting reputation-based trust for peer-to-peer electronic communities," *IEEE transactions on Knowledge and Data Engineering,* vol. 16, no. 7, pp. 843-857, 2004.
[13] X. Liu, A. Datta, K. Rzadca, and E.-P. Lim, "StereoTrust: a group based personalized trust model," in *Proceedings of the 18th ACM conference on Information and knowledge management*, Hong Kong, China, 2009, pp. 7-16.
[14] L. Atzori, A. Iera, and G. Morabito, "From "smart objects" to "social objects": The next evolutionary step of the internet of things," *IEEE Communications Magazine,* vol. 52, no. 1, pp. 97-105, 2014.
[15] S. Nepal, W. Sherchan, and C. Paris, "Strust: A trust model for social networks," in *IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2011, pp. 841-846.
[16] Y. Hu, D. Wang, H. Zhong, and F. Wu, "SocialTrust: Enabling long-term social cooperation in peer-to-peer services," *Springer Peer-to-Peer Networking and Applications,* vol. 7, no. 4, pp. 525-538, 2014.
[17] M. Nitti, R. Girau, L. Atzori, A. Lera, and G. Morabito, "A Subjective Model for Trustworthiness Evaluation in the Social Internet of Things," in *IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC*, Australia, 2013, pp. 18-23.
[18] Z. Liang, and W. Shi, "Enforcing cooperative resource sharing in untrusted P2P computing environments," *Mob. Netw. Appl.,* vol. 10, no. 6, pp. 971-983, 2005.
[19] Y. Wang, and J. Vassileva, "Bayesian Network-Based Trust Model," in *IEEE International Conference on Web Intelligence,(WI'03).* 2003, pp. 372.
[20] W. Li, W. Meng, L.-F. Kwok, and H. Horace, "Enhancing collaborative intrusion detection networks against insider attacks using supervised intrusion sensitivity-based trust management model," *Journal of Network and Computer Applications,* vol. 77, pp. 135-145, 2017.
[21] A. Bolster, and A. Marshall, "Analytical metric weight generation for multi-domain trust in autonomous underwater MANETs," in *IEEE Third Underwater Communications and Networking Conference (UComms)*, Lerici, Italy, 2016, pp. 1-5.
[22] Y. W. Lee, D. M. Strong, B. K. Kahn, and R. Y. Wang, "AIMQ: a methodology for information quality assessment," *Information & management,* vol. 40, no. 2, pp. 133-146, 2002.
[23] L. L. Pipino, Y. W. Lee, and R. Y. Wang, "Data quality assessment," *Commun. ACM,* vol. 45, no. 4, pp. 211-218, 2002.
[24] B. Heinrich, M. Kaiser, and M. Klier, "How to measure data quality? A metric-based approach," ICIS 2007 Proceedings, p. 108, 2007.
[25] N. Askham, D. Cook, M. Doyle, H. Fereday, M. Gibson, U. Landbeck, R. Lee, C. Maynard, G. Palmer, and J. Schwarzenbach, *The six primary dimensions for data quality assessment*, Technical report, DAMA UK Working Group, 2013.
[26] S. Mazilu, M. Teler, and C. Dobre, "Securing vehicular networks based on data-trust computation," pp. 51-58.
[27] M. Raya, P. Papadimitratos, V. D. Gligor, and J.-P. Hubaux, "On data-centric trust establishment in ephemeral ad hoc networks. " pp. 1238-1246.
[28] P. Borzymek, M. Sydow, and A. Wierzbicki, "Enriching trust prediction model in social network with user rating similarity. "in International Conference on Computational Aspects of Social Networks, CASON'09, , pp. 40-47, 2009.
[29] N. Korovaiko, and A. Thomo, "Trust prediction from user-item ratings," *Social Netw. Analys. Mining,* vol. 3, no. 3, pp. 749-759, 2013.
[30] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks." in proceedings of the 19th international conference on World wide web, pp. 981-990, 2010.
[31] C. Castelfranchi, and R. Falcone, "Principles of trust for MAS: Cognitive anatomy, social importance, and quantification," in *Multi Agent Systems, 1998. Proceedings. International Conference on*, 1998, pp. 72-79.
[32] P. R. Benson, "ISO 8000 Data Quality: *The Fundamentals Part 1*," Real-World Decision Support (RWDS) Journal 3, no. 4, 2009.
[33] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, Las Vegas, Nevada, USA, 2008, pp. 426-434.
[34] Y.-x. Yuan, "Step-sizes for the gradient method," *AMS IP Studies in Advanced Mathematics,* vol. 42, no. 2, pp. 785, 2008.
[35] A. Jian, G. Xiaolin, Y. Jianwei, S. Yu, and H. Xin, "Mobile crowd sensing for internet of things: A credible crowdsourcing model in mobile-sense service," in *Multimedia Big Data (BigMM), 2015 IEEE International Conference on*, 2015, pp. 92-99.
[36] Twin Oaks Computing Inc. "CoreDX Distributed Publish-Subscribe System," [Online] Available: http://www. twinoakscomputing.com/coredx/develop.