Model-Based Trust Assessment for Internet of Things Networks

Stephen Adams, Peter A. Beling
Systems and Information Engineering
School of Engineering and Applied Sciences
University of Virginia
Charlottesville, VA 22903

Steven Greenspan
CA Technologies
200 Princeton S. Corp Center
Ewing, NJ 08628

Maria Velez-Rojas, Serge Mankovski
CA Technologies
3965 Freedom Circle
Santa Clara, CA 95054

***Abstract*—Trust in data collected by and passing through Internt of Things (IoT) networks is paramount. The quality of decisions made based on this collected data is highly dependent upon the accuracy of the data. Currently, most trust assessment methodologies assume that collected data follows a stationary Gaussian distribution. Often, a trust score is estimated based upon the deviation from this distribution. However, the underlying state of a system monitored by an IoT network can change over time, and the data collected from the network may not consistently follow a Gaussian distribution. Further, faults that occur within the estimated Gaussian distribution may go undetected. In this study, we present a model-based trust estimation system that allows for concept drift or distributions that can change over time. The presented methodology uses data-driven models to estimate the value of the data produced by a sensor using the data produced by the other sensors in the network. We assume that an untrustworthy piece of data falls in the tails of the residual distribution, and we use this concept to assign a trust score. The method is evaluated on a smart home data set consisting of temperature, humidity, and energy sensors.**

I. Introduction

Data is being collected at an ever increasing rate, and technological advances are allowing users to store and analyze the vast amounts of data present in modern society. The sources of data are diverse and include sensor networks, smart phones, cyber-physcial systems, and social networks. End users extract information from collected data and ultimately wish to make decisions utilizing this information. One common interest to all of these domains is the need to evaluate the trustworthiness of the data being collected and supplied to end users. Corrupted data, either intentionally corrupted or corrupted through natural causes, supplied to an end user or decision maker will degrade the performance of the system and could lead to incorrect or even fatal decisions. For example, untrustworthy IoT data in a smart-city setting [1] could lead a decision maker to deploy city resources, such as fire fighters and ambulances, to incorrect or unnecessary locations. This mismanagement of resources could cause waste and resources not being available for real events such as fires or medical emergencies.

Internet of things (IoT) networks are growing in popularity and becoming a larger source of data. Privacy, security, and trust play a prominent role in any IoT deployment [2], [3], [4], [5]. Trust in IoT networks can be broken down into two concepts: (1) trust in the interactions between entities in the network, and (2) trust in the network itself [4]. This study focuses on evaluating the trust of an IoT network, and in particular, mechanisms for a user to evaluate the trustworthiness of the data produced by an IoT network.

Trust is a difficult concept to define, and researchers often define a narrow definition of trust that reflects their empirical study [6]. In a network, trust can be thought of as the belief that an entity will perform an action given a particular circumstance [7]. This definition directly applies to peer-to-peer (P2P) networks where entities are interacting with each other. However, this definition does not align with a network that solely provides information or data. We amend this definition to apply to IoT networks that produce data:

Definition 1. *Data Trustworthiness in IoT Networks* Trust in data in IoT networks is the subjective probability that data observed by a user is consistent with the data at the source.

Given this definition, we wish to construct models that can accurately assess the trust a user can place in a piece of data collected or produced by an IoT network.

The concept of trust has been applied to numerous types of systems and networks. P2P networks link users directly and allow them to share files. These networks are not only a security risk but also require that user be evaluated in terms of competence or quality. Bayesian networks have been used to estimate the trustworthiness of file providers based on several attribute including download speed and file quality [8], [9], [10]. Bearly and Kumar [11] develop a general framework for trust in a P2P network where entities are included in the *trust* network if they are reliable. Ad hoc networks lack infrastructure and have dynamic topologies. Trust models for these types of networks must account for these characteristics [12], [13]. In both of these networks, the trust model estimates the reputation of the entity. It is assumed that an entity with a good reputation will provide quality data, but the trust of the data is not explicitly estimated.

Wireless sensor networks (WSNs) are composed of small autonomous nodes. Due to a self-contained power supply, each node has a limited communication and computational capability. Jiang *et al* [14] develop an efficient trust framework

for WSNs. They draw a distinction between direct and indirect trust, where the former is calculated based on interaction and the latter is calculated based on a third-party assessment. Hur *et al* [15] create a trust model that uses data to distinguish between "illegal" and "legal" nodes. The basic premise of this model is to collect multiple redundant data source and compare them for consistency. Feng *et al* [16] offer a trust evaluation algorithm for WSNs that incorporates a number of trust factors into the final trust calculation. Won and Bertino [17] assume that data similarity is correlated with physical distance between nodes.

Yan, Zhang, and Vasilakos, [18] provide a survey of trust management in IoT networks and cover several topics including trust properties, trust evaluation, trust frameworks, and the perception of data trust. Sicari *et al.* [19] survey the literature on security, privacy, and trust for IoT networks. WSNs and IoT networks both require some form of data trust evaluation, however most of the existing methods rely on the behavior of nodes or the ability to evaluate the interactions between nodes and do not consider the data itself. Javed and Wolf [20] develop a method for estimating the trustworthiness of data in an IoT network based on the physical distance between nodes and outlier detection. A pattern-wise trust assessment [21] allows for nodes to be grouped into neighborhoods based on different states of the world.

Several trust models utilize data provenance to generate trust scores. The W7 model [22] conceptualizes provenance as a combination of several factors: what, when, where, why, how, who, and which. Dai *et al.* [23] make the case for using provenance to estimate the trust of data and the data sources. They break provenance, or the information about the data used to estimate the trust score, into four attributes: data similarity, data conflict, path similarity, and data deduction. In [24] and [25], Lim, Moon, and Bertino present a model for evaluating the trustworthiness of data and nodes in a network for streaming data. The proposed cyclic framework uses models for the data, provenance, and similarity measures to estimate trust scores. Wang, Govindan, and Mohapatra [26] estimate a provenance-based notion of trust using path similarity and information similarity.

Most of the previously mentioned methods rely on a similarity metric that compares past data with newly collected data and then assigns a trust score. Often, the similarity metric assumes that the data follows a stationary Gaussian distribution. In these cases, data that may be not be trustworthy could fool the trust model if it is "close" to the mean of the estimated Gaussian distrubtion. Systems can also evolve over time, which can cause new trustworthy data to appear as untrustworthy data. To combat these issues, we propose a model-based trust assessment method that estimates a value for the data, and then evaluates the trustworthiness of the data based on the residual between the estimated value and the collected value. The outlier-based method in [20] has a similar concept but we build on that work in 2 key areas. First, Javed and Wolf specifically recommend using a linear regression model for predicting the future value of the data. We generalize and recommend exploring several types of data-driven models and selecting the model that is best for the application. Second, Javed and Wolf do not specify how to determine if an observation is an outlier. In the proposed method, we provide a method for detecting an outlier based on the residuals and convert this value into a trust score.

The primary contribution of this study is the use of residuals from a data-driven model to estimate a trust score for individual pieces of data. In Section II, the proposed model-based method for estimating trust is outlined. Section III provides numerical experiments on an IoT dataset. In Section IV, we provide our conclusions on the proposed method.

II. Method

Let $X_i(t)$ be the value of the data from node i in an IoT network at time t for $i = 1...I$ and $t = 1...T$. Further, let $\mathbb{X}(t) = \{X_1(t), ..., X_I(t)\}$, the set of all the data from the nodes in the IoT network at t. In the proposed method, we assume that there is a relationship between all the nodes in the network that can be characterized using a data-driven model. More specifically, we assume that $X_i(t)$ can be estimated using the data from all the other nodes in the network represented by $\bar{X}(t) = \{X_j(t) \in \mathbb{X}(t) | 1 \leq j \leq I, j \neq i\}$. Let $f_i(\cdot)$ be a model that estimates $X_i(t)$, where

$$\hat{X}_i(t) = f_i(\bar{X}(t)). \tag{1}$$

The model $f_i(\cdot)$ is estimated or learned during a training process from training data $\mathcal{X}$ procured sometime before t. Let T' be the number of observations in $\mathcal{X}$ and $\tau = 1...T'$. The proposed model-based trust assessment method compares $\hat{X}_i(t)$ with $X_i(t)$ and then estimates a trust score. Let $R_i(t)$ be the residual

$$R_i(t) = \hat{X}_i(t) - X_i(t). \tag{2}$$

Residuals follow a student's t-distribution, and the parameters of the residual distribution $\{\mu, \sigma, \nu\}$ are estimated from the residuals of $\mathcal{X}$. Inspired by the work in [24], a trust score $S_i(t)$ is estimated using

$$S_i(t) = 2 \int_{\bar{R}_i(t)}^{\infty} p(x|\nu)dx, \quad \text{if } \bar{R}_i(t) < 0, \tag{3}$$

and

$$S_i(t) = 2 \int_{0}^{\bar{R}_i(t)} p(x|\nu)dx, \quad \text{if } \bar{R}_i(t) \geq 0, \tag{4}$$

where $\bar{R}_i(t)$ is the standardized residual, and $p(x|\nu)$ is the probability density function of the standardized student's t-distribution. Note that μ and σ are omitted from $p(x|\nu)$ because $\mu = 0$ and $\sigma = 1$ for the standardized distribution. The primary difference between Equations 3 and 4 and the intermediate trust score in [24] is that we estimate trust based on the residual while Lim, Moon, and Bertino estimate the intermediate trust score using the raw data. When standardizing the residuals, α can be added to σ to increases the standard

deviation of the estimated residual distribution. Under the estimated distribution, some trustworthy observations would naturally occur in the tails of the distribution. By increasing the variance of the estimated distribution, the likelihood of trustworthy observations assigned low trust scores is decreased. Algorithm 1 displays the entire model-based trust scoring algorithm.

Algorithm 1: Model-Based Trust Scoring Algorithm

Result: Trust scores for all nodes $i = 1...I$ at all time points $t = 1...T$

Build models $f_i(\cdot)\ \forall i$ using $\mathcal{X}$
Calculate $R'_i(\tau)$ for $\tau = 1...T'$ and $i = 1...I$
Estimate μ_i, σ_i, and ν_i from R'_i
Increase standard deviation of estimated distribution $\sigma = \sigma + \alpha$
for $t = 1...T$ **do**
  for $i = 1...I$ **do**
    $\hat{X}_i(t) = f_i(\bar{X}(t))$
    $R_i(t) = \hat{X}_i(t) - X_i(t)$
    $\bar{R}_i(t) = (R_i(t) - \mu)/\sigma$
    if $\bar{R}_i(t) < 0$ **then**
      $S_i(t) = 2\int_{\bar{R}_i(t)}^{\infty} p(x|\nu)dx$
    else
      $S_i(t) = 2\int_{0}^{\bar{R}_i(t)} p(x|\nu)dx$
    end
  end
end

III. Experiments

The model-based trust scoring method outlined in Algoirthm 1 is demonstrated on a smart home data set [27] publicly available on the UCI machine learning repository[1]. Temperature, relative humidity, and energy consumption are collected from inside the home. Weather data from the local airport is also available. The data set contains over 19,000 observations and was collected every 10 minutes from January 11, 2016 to May 27, 2016. Fig. 1 displays the relative humidity data from sensor RH 1. The numerical experiments are limited to this sensor, but it could be easily applied to any of the sensors in the dataset. The first month of the data is used for training, and the trust score is evaluated for the rest of the data.

First, we will point out a limitation of standard trust algorithms when trying to estimate the trust score for RH 1. The reputation-based framework for assessing the integrity of sensors presented in [28] relies heavily on outlier detection algorithms, and the authors recommend the local outlier factor (LOF) method [29]. The LOF algorithm is applied to the RH 1 data, and Fig. 2 displays the results. The algorithm assigns non-outliers a value of 1, and outliers the value of 2. Given the training data, the algorithm identifies several outliers

[1]https://archive.ics.uci.edu/ml/datasets/Appliances+energy+prediction

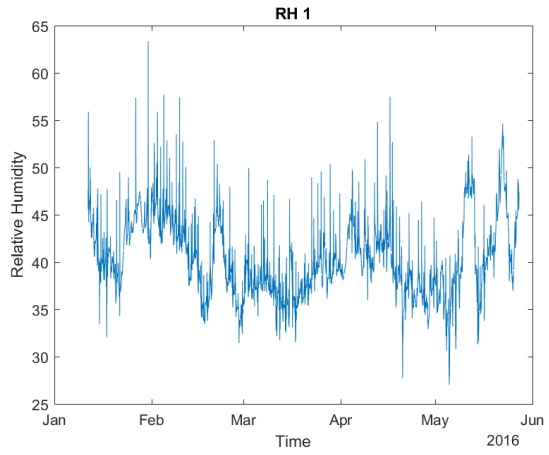

Fig. 1. Relative humidity data from sensor RH 1.

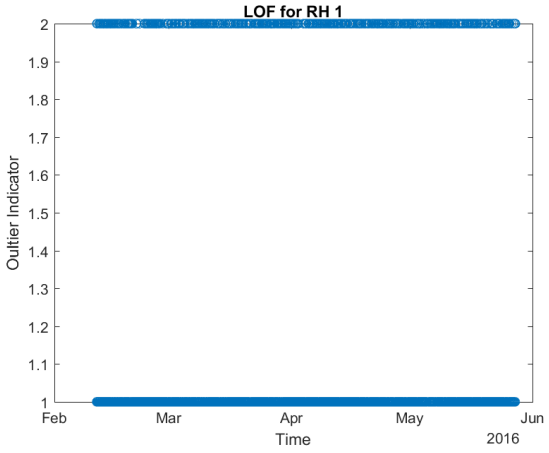

Fig. 2. LOF algorithm applied to RH 1. A 1 indicates a non-outlier, and a 2 indicates an outlier as predicted by the algorithm.

throughout the testing dataset even though these are legitimate data values. This will significantly affect the results of the reputation-based system.

The authors of [28] present several types of faults that can be injected into a data stream to represent untrustworthy data. The offset fault, displayed in Fig. 3, adds an offset value m to the true data value. The variance fault, displayed in Fig. 4, adds a value sampled from a Gaussian distribution with mean 0 and variance v to the true data value. The stuck fault, displayed in Fig. 5, replaces the true data value with a single value s. The LOF algorithm is applied to data with the stuck fault, and the outlier detection results are displayed in Fig. 6. The LOF algorithm completely misses the fault.

The presented model-based method is applied to the RH 1 data stream. First, a model that predicts the relative humidity from RH 1 must be selected. The inputs into the model are the other relative humidity sensors in the data set, the temperature sensors, the energy sensors, and the local weather data. Several models were tested including linear regression, random forest regression, gradient boosted machines, and

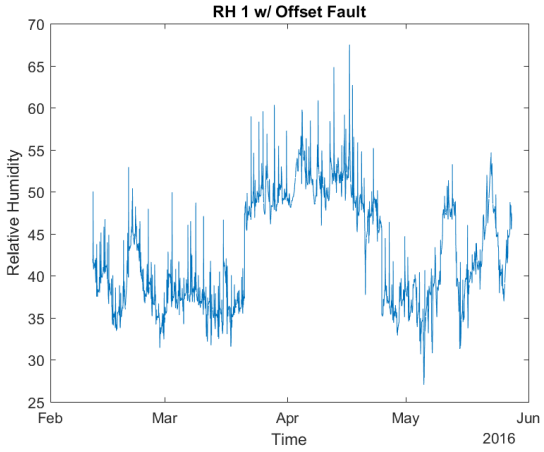

Fig. 3. RH 1 with an offset fault injected from late March to late April.

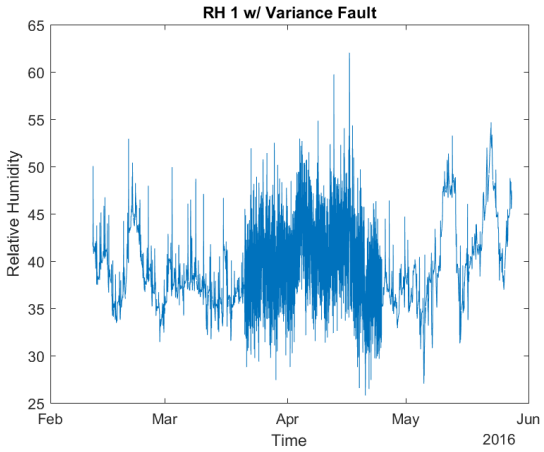

Fig. 4. RH 1 with a variance fault injected from late March to late April.

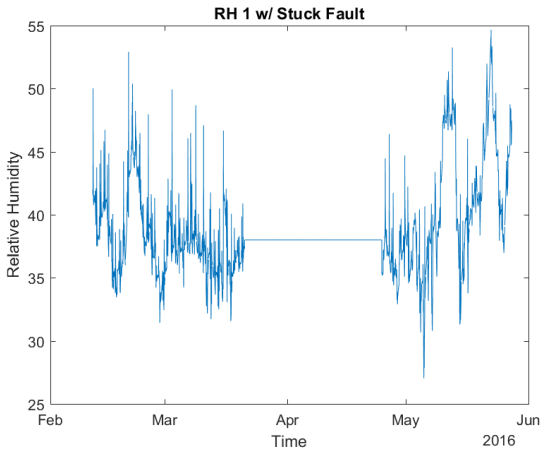

Fig. 5. RH 1 with a stuck fault injected from late March to late April.

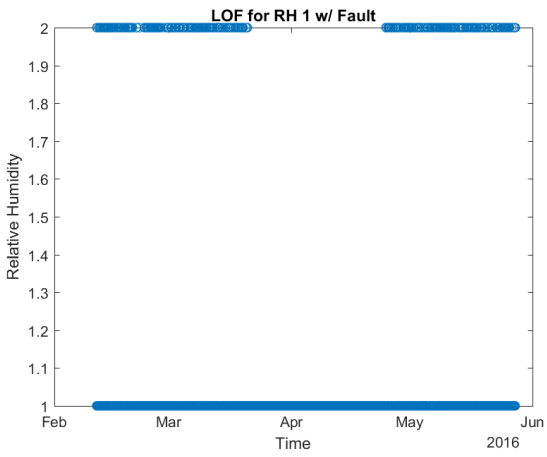

Fig. 6. LOF algorithm applied to RH 1 with a stuck fault. A 1 indicates a non-outlier, and a 2 indicates an outlier as predicted by the algorithm.

TABLE I
ROOT MEAN SQUARE ERROR (RMSE)

Model	Training RMSE	Testing RMSE
Linear Regression	1.1	1.6
Random Forest	0.6	3.0
Gradient Boosted Machines	0.5	5.3
Multilayer Perceptrons	3.0	4.6

multilayer perceptrons. Table I displays the root mean square error for the training and testing sets. The training set error was estimated using 10-fold cross validation. The gradient boosted machine has the lowest training error but does not generalize well to the test set. Linear regression was chosen because of the low error on the test set.

First, we test the presented method on the testing data from RH 1 under normal conditions and an $\alpha = 3$ (Fig. 7). The trust scores are predominantly above 0.4. Fig. 8, Fig. 9, and Fig. 10 display the trust scores for the offset, variance, and stuck faults, respectively. The algorithm clearly identifies the offset fault and the stuck fault. It appears that the trust score decreases in the region of the variance fault.

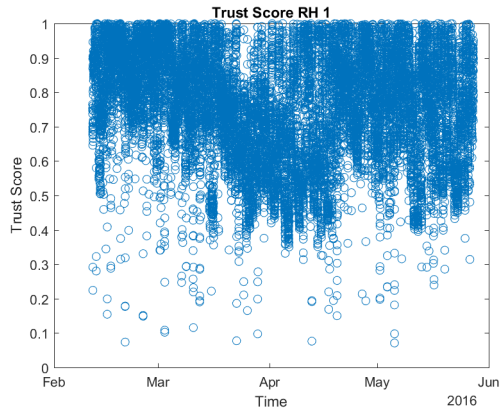

Fig. 7. Model-based trust scoring algorithm applied to RH 1.

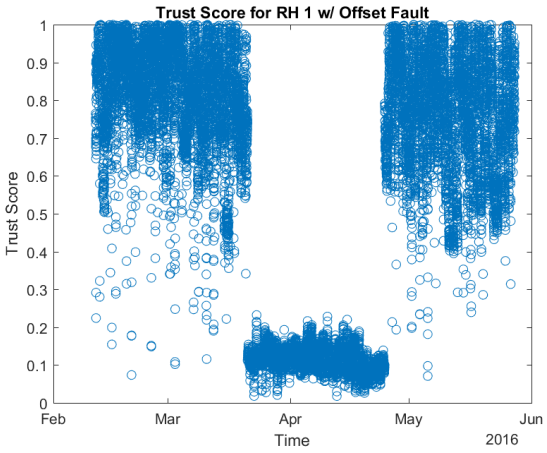

Fig. 8. Model-based trust scoring algorithm applied to RH 1 w/ an offset fault.

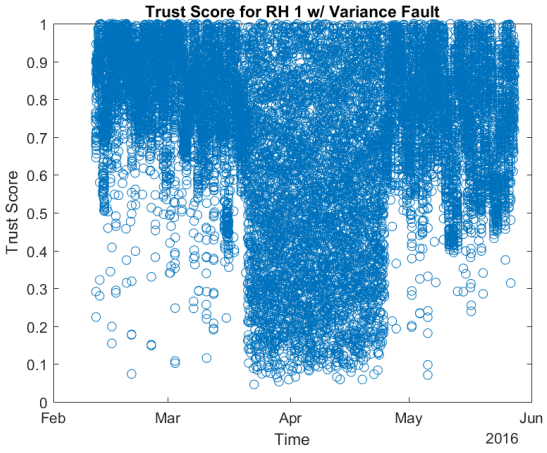

Fig. 9. Model-based trust scoring algorithm applied to RH 1 w/ a variance fault.

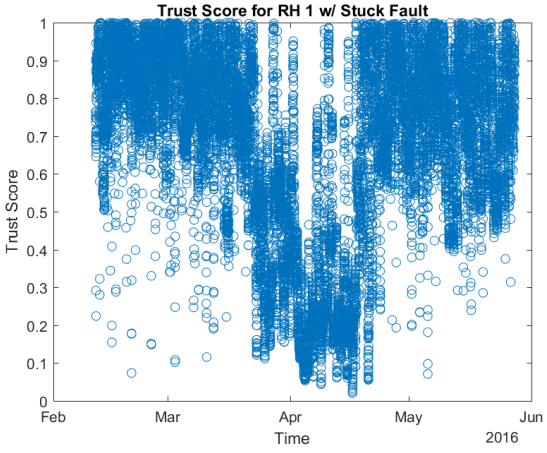

Fig. 10. Model-based trust scoring algorithm applied to RH 1 w/ a stuck fault.

TABLE II
MODEL-BASED SENSITIVITY AND SPECIFICITY

Fault	Sensitivity	Specificity
Offset	1.00	0.99
Variance	0.32	0.99
Stuck	0.51	0.99

TABLE III
LOF SENSITIVITY AND SPECIFICITY

Fault	Sensitivity	Specificity
Offset	0.06	0.89
Variance	0.21	0.89
Stuck	0	0.89

To further evaluate the proposed method, individual data points are classified as either trustworthy or untrustworthy using a threshold of 0.4, i.e. data points with a trust score below 0.4 are classified as untrustworthy. For this analysis, we consider the faulty data point to be a positive outcome. Sensitivity and specificity for each fault are displayed in Table II. The proposed method has a specificity of 0.99 for all faults meaning that trustworthy data points are rarely assigned trust scores below the threshold. The offset fault has a sensitivity of 1 meaning that all of the untrustworthy data points have a trust score below the threshold. The method identifies just over half the untrustworthy data points for the stuck fault, but it only identifies a third of untrustworthy data points for the variance fault. As a comparison, Table III displays the sensitivity and specificity for the LOF outlier detection algorithm. The model-based method outperforms the LOF outlier detection method on both performance metrics.

IV. CONCLUSION

A model-based trust scoring algorithm is presented and tested on a smart-home IoT data set. The method builds upon previous work by generalizing to a data-driven model, specifying a method for detecting outliers, and converting the outlier detection to a trust score. It is demonstrated that other common outlier detection methods, specifically the LOF method, have trouble on this dataset when the first month of the data is used for training. Outliers are often detected for true values and the method misses faults that lie within the bounds of the training data. The presented model-based method can detect the areas in the data where three common faults are injected.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. CNS: 1650512. This work was conducted in the NSF IUCRC Center of Visual and Decision Dynamics, through the sponsorship and guidance of CA Technologies. We thank the folks on the team who choose not be authors for their thoughts on this paper and the overall project.

REFERENCES

[1] S. P. Mohanty, U. Choppali, and E. Kougianos, "Everything you wanted to know about smart cities: The internet of things is the backbone," *IEEE Consumer Electronics Magazine*, vol. 5, no. 3, pp. 60–70, 2016.

[2] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of things: Vision, applications and research challenges," *Ad hoc networks*, vol. 10, no. 7, pp. 1497–1516, 2012.

[3] D. Puthal, N. Malik, S. P. Mohanty, E. Kougianos, and C. Yang, "The blockchain as a decentralized security framework," *IEEE Consumer Electronics Magazine*, vol. 7, no. 2, pp. 18–21, 2018.

[4] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.

[5] R. H. Weber, "Internet of Things–new security and privacy challenges," *Computer law & security review*, vol. 26, no. 1, pp. 23–30, 2010.

[6] D. H. McKnight and N. L. Chervany, "What is trust? a conceptual analysis and an interdisciplinary model," *AMCIS 2000 Proceedings*, p. 382, 2000.

[7] G. Dogan and T. Brown, "A survey of provenance leveraged trust in wireless sensor networks," *Computer Engineering and Intelligent Systems*, vol. 5, no. 2, pp. 1–11, 2014.

[8] Y. Wang and J. Vassileva, "Bayesian network-based trust model," in *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*. IEEE, 2003, pp. 372–378.

[9] ——, "Bayesian network trust model in peer-to-peer networks," in *International Workshop on Agents and P2P Computing*. Springer, 2003, pp. 23–34.

[10] ——, "Trust and reputation model in peer-to-peer networks," in *Peer-to-Peer Computing, 2003.(P2P 2003). Proceedings. Third International Conference on*. IEEE, 2003, pp. 150–157.

[11] T. Bearly and V. Kumar, "Expanding trust beyond reputation in peer-to-peer systems," in *Database and Expert Systems Applications, 2004. Proceedings. 15th International Workshop on*. IEEE, 2004, pp. 966–970.

[12] Z. Liu, A. W. Joy, and R. A. Thompson, "A dynamic trust model for mobile ad hoc networks," in *Distributed Computing Systems, 2004. FTDCS 2004. Proceedings. 10th IEEE International Workshop on Future Trends of*. IEEE, 2004, pp. 80–85.

[13] Y. Rebahi, V. E. Mujica-V, and D. Sisalem, "A reputation-based trust mechanism for ad hoc networks," in *Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on*. IEEE, 2005, pp. 37–42.

[14] J. Jiang, G. Han, F. Wang, L. Shu, and M. Guizani, "An efficient distributed trust model for wireless sensor networks," *IEEE transactions on parallel and distributed systems*, vol. 26, no. 5, pp. 1228–1237, 2015.

[15] J. Hur, Y. Lee, H. Youn, D. Choi, and S. Jin, "Trust evaluation model for wireless sensor networks," in *Advanced Communication Technology, 2005, ICACT 2005. The 7th International Conference on*, vol. 1. IEEE, 2005, pp. 491–496.

[16] R. Feng, X. Xu, X. Zhou, and J. Wan, "A trust evaluation algorithm for wireless sensor networks based on node behaviors and ds evidence theory," *Sensors*, vol. 11, no. 2, pp. 1345–1360, 2011.

[17] J. Won and E. Bertino, "Distance-based trustworthiness assessment for sensors in wireless sensor networks," in *International Conference on Network and System Security*. Springer, 2015, pp. 18–31.

[18] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for Internet of Things," *Journal of network and computer applications*, vol. 42, pp. 120–134, 2014.

[19] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Computer Networks*, vol. 76, pp. 146–164, 2015.

[20] N. Javed and T. Wolf, "Automated sensor verification using outlier detection in the internet of things," in *Distributed Computing Systems Workshops (ICDCSW), 2012 32nd International Conference on*. IEEE, 2012, pp. 291–296.

[21] R. Gwadera, M. Riahi, and K. Aberer, "Pattern-wise trust assessment of sensor data," in *Mobile Data Management (MDM), 2014 IEEE 15th International Conference on*, vol. 1. IEEE, 2014, pp. 127–136.

[22] S. Ram and J. Liu, "A new perspective on semantics of data provenance," in *Proceedings of the First International Conference on Semantic Web in Provenance Management-Volume 526*. CEUR-WS. org, 2009, pp. 35–40.

[23] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *Workshop on Secure Data Management*. Springer, 2008, pp. 82–98.

[24] H.-S. Lim, Y.-S. Moon, and E. Bertino, "Provenance-based trustworthiness assessment in sensor networks," in *Proceedings of the Seventh International Workshop on Data Management for Sensor Networks*. ACM, 2010, pp. 2–7.

[25] ——, "Assessing the trustworthiness of streaming data," *Technical Report TR 2010–09, CERIAS, West Lafayette*, 2010.

[26] X. Wang, K. Govindan, and P. Mohapatra, "Provenance-based information trustworthiness evaluation in multi-hop networks," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*. IEEE, 2010, pp. 1–5.

[27] L. M. Candanedo, V. Feldheim, and D. Deramaix, "Data driven prediction models of energy use of appliances in a low-energy house," *Energy and Buildings*, vol. 140, pp. 81–97, 2017.

[28] S. Ganeriwal, L. K. Balzano, and M. B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 4, no. 3, p. 15, 2008.

[29] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *ACM sigmod record*, vol. 29, no. 2. ACM, 2000, pp. 93–104.