Provenance Based Information Trustworthiness Evaluation in Multi-hop Networks

Xinlei (Oscar) Wang, Kannan Govindan and Prasant Mohapatra
Department of Computer Science
University of California, Davis, CA 95616
Email: {xlwang, kgovindan, pmohapatra}@ucdavis.edu

***Abstract*—In this paper, we present a trust model to evaluate the trustworthiness of information as well as the information publishing nodes based on the information provenance. We consider two factors in evaluating the provenance based information trust: Path similarity and Information similarity. In multihop networks, information can flow through multiple hops and in multiple paths. We model the similarity factor between different paths which deliver information about the same event. We also model the similarity between two information items delivered through different paths about the same event. Both path and information similarity factors are considered in determining the trust on the information. This information trust indeed used as a feedback factor to adaptively adjust trust of the nodes in the network. Detailed analysis of the proposed approach is presented along with simulation results for validation.**

I. Introduction

In a dynamic network environment, a mechanism of trust and reputation evaluation is an indispensable component to enhance the security of the entire network. In such an network environment, information transmissions and sharing are some of the essential activities, and thus the quality of information is crucial to the end users especially to the decision makers. By analyzing the trustworthiness of information received from different network nodes or entities, the decision makers can evaluate the quality of information received from them and make the right decisions. In a multi-hop network, information is generated from a source node, e.g., a sensor. It then may need to go through a series of other intermediate nodes before reaching its destination, i.e., the end user.

Lots of work has been done on the protection from data tampering, e.g., digital signature techniques, to ensure data integrity when the information routed through multiple nodes. However, they do not address the problem of information trustworthiness. Untrustworthy information may be introduced because of two different reasons: *unintentional errors* and *intentional misbehaviour* [1]. Unintentional errors are caused by malfunction of the hardware (e.g., broken or obstructed sensors), mispositioning of the node or exhausted batteries. Intentional misbehaviour is caused by malicious attackers, providing false data on purpose through a compromised node. In order to assess the trustworthiness of information, we need to consider not only the trustworthiness of its direct sender but also the provenance of the information. We define the trustworthiness of information items and nodes as well as information provenance as follows.

DEFINITION 1: **Trustworthiness of Information Items (Trust):** The trustworthiness of an information item i, denoted as $T(i)$, is the probability of i being true. In this paper, we use the term "trust" to represent the trustworthiness of information items.

DEFINITION 2: **Trustworthiness of Nodes (Reputation):** The reputation of a node N, denoted as R_N, is the synthesized probability that N annotates correct trust value on the information items it owns, perceived by all end users which have been communicating with node N. We use the term "reputation" to denote the trust of a node.

DEFINITION 3: **Information Provenance:** The provenance details about the information contains history of the information starting from its creation. It has details about the owner of the information and details about the various nodes which have passed/processed the information before it reaches the destination.

We propose a three-step approach to evaluate the trustworthiness of the information and the information providers based on the provenance details. We first estimate the trustworthiness of information based on the trustworthiness of its provider, then we further assess the trustworthiness of this information based on similarity of information received from multiple paths and the correlation between the paths through which the information items are delivered. Finally, we adapt the trustworthiness of the information provider in a feedback manner using the calculated new information trust value.

The rest of the paper is organized as follows. Section II summaries some of the related works in the field of reputation techniques and provenance based information quality assessment. Section III introduces the fundamental framework and concepts of the provenance-based information trustworthiness evaluation. Section IV presents our proposed information trust model and the detailed computation framework. Analysis and simulation results are presented in Section V. Finally, Section VI concludes the paper and outlines our future research work.

II. Related works

Reputation techniques for evaluating the trust of different entities in a network environment have been widely studied, e.g., [2]–[8]. These techniques are important to detect any network nodes or group of nodes that behave maliciously and thus enhance the security of the overall network. However, they do not address the question of whether we can trust certain information which reaches a decision maker through a series of different nodes. Besides, there exist situations that an intermediate node generates new information based on inputs from other nodes. How much can we trust this kind of infused/newly-added information? To address these questions, we need to make use of the provenance associated with the information. However, very limited work has considered modeling and analysis approaches to assess the trustworthiness of information based on provenance.

A related work is an agent-based approach proposed by Yu et al. [9], in which a computation model is presented to calculate the trustworthiness of information in a dynamic information sharing environments using the framework of Desmpster-Shafer theory. The problem of this approach is that Dempster-Shafer theory cannot be used to correctly capture information conflicts and Dempster's rule of combination can only merge independent evidences whereas in a dynamic network environment, correlation between information items could be very common, e.g., information reached at the end user node from different paths actually originated from a same node or has been processed by some common nodes. A data provenance trust model which estimates the level of trustworthiness of both information and information providers is presented in [10]. Four aspects that affect the trustworthiness of the data have been taken into account to build this trust model, which are (a) data similarity, (b) path similarity, (c) data conflict and (d) data deduction. This approach also has certain drawbacks when applying to network situations. First, the data deduction trust computation is too simple and fixed which may lead to inaccurate results. Second, an intermediate node can only publish what it thinks is fully trustworthy in order to prevent its own trust from getting decreased, whereas there could be a lot of situations that some information is not 100% trustworthy but still valuable to the end users.

In contrast to the previous work, our approach is unique by taking both information and path correlation into consideration and also evaluating the reputation of nodes in a feedback manner based. In addition, our model is independent of techniques used by intermediate nodes to derive/infuse new information, hence they can have many different ways to do so based on different purposes and the end users need not to have any knowledge about it. Besides, intermediate nodes are allowed to process and send valuable data that are not necessarily fully trustworthy in a regulated way without harming their own reputations.

III. Fundamental Framework

In this paper, we consider a scenario in which we have three types of entities in the network: source nodes, intermediate nodes and end users. Source nodes could be sensor nodes that generate new information about certain events which is then relayed by the intermediate nodes to the end users. End users receives the information and evaluate the trustworthiness of this information then make their decisions on further actions. Note that the further actions here could be just modifying the information and forwarding it to other nodes. In this case, the end user will be considered as an intermediate node by the receiver of the modified information. Therefore, intermediate nodes not only can just "pass" the information but also can process the information based on their own judgements. In addition, they can also generate some new information based on information it receives from different nodes.

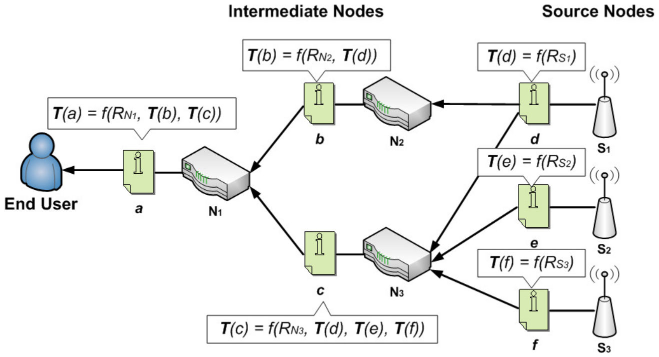

Fig. 1. An Example Network Scenario for Provenance-based Information Trust Evaluation

Here we consider information in the form of information items. The information items are statements about certain external events in the network environment, such as sensor reports. Each information item has an owner, i.e., the source node or intermediate node which published the information item. When an intermediate node N processes or merges some old information, we consider the resulting information item generated by N as a completely new information item and thus its owner is N rather than the original source node(s). Each information item consists of information meta data and information payload. The information meta data in turn contains the provenance of the information item. The trust of the information item is function of provenance. For instance in Fig. 1 the trust of information b is a function $f(.)$ of the reputation value of node N_2 (R_{N_2}) and also the trust of information item d ($T(d)$).

As shown in the Fig. 1, the user receives information item a from node N_1. By looking into its provenance, the user knows node N_1 derived information a based on information items b and c. b's provenance indicates that it is sent by the intermediate node N_2, either "passed" by N_2 (so that b and d are the same information item and the owner is S_1) or processed by N_2 (so that b is a new information item and the owner is N_2). c's provenance indicates that it is a combination of information items d, e and f, and the owner is node N_3. Finally, d, e and f are owned by source nodes S_1, S_2 and S_3 respectively.

The primary interest of our investigation is the trustworthiness of the information items and the information providers (source and intermediate nodes). In our network model, every node has a single reputation value available globally that reflects the opinions on its trustworthiness of all other nodes whom the node has been communicating with. Any new nodes entering the network will have a default reputation value of 0.5. A node N always annotates a trust value of the information items it owns with its own signature in the meta data so that the end users can adjust node N's reputation value based on the correctness of the information trust reported by N. We can gain three advantages by doing this:

1) Intermediate nodes can send valuable information items which are not necessarily fully trustworthy to downstream nodes without harming their own reputation values. In other words, even when the trustworthiness of an information item is very minimal (≈ 0), a node can still send it without harming its own reputation value as long as it annotates the right trust value on the information item.
2) Different nodes may have different ways of processing or merging information items and the associated trust value based on different purposes and the end users need not know about it.
3) False information with incorrect annotated trust value will decrease the owner's reputation only, and thus nodes only take responsibility for the information items they own.

Digital signature techniques have been developed maturely, it is easy to achieve and ensure that every node sends out data with signatures. We assume no unauthorized data tampering without signatures in our system. In addition, our current model focuses on security concerns caused by individual compromised nodes, we assume no collusion attacks.

IV. The Provenance-based Trust Model

In this section, we present our information provenance trust model to assign trust values to information items and adjust reputation values of nodes in the network. Trust and reputation values range from 0 to 1. A value of 1 means completely trustworthy, 0 means the opposite and 0.5 means completely uncertain about the trustworthiness.

In order to assign trust values to information items and adjust reputation values of network nodes, we divide our trust model into the following three steps:

- Initial trust computation
- Information trust adjustment
- Reputation feedback.

In the rest of this section, we present the details of the these steps.

A. *Initial Trust Computation*

In the first step, we consider the following question: when a node (end user) receives an information item i whose owner is some node N, how can the user decide how much it can trust i, given that the user knows nothing that can help with its assessment except for the provenance of information item i and the reputation value of node N. We need to note that the information item i has a trust value reported by N ($T_N(i)$) in its meta data. It is intuitive to see that the higher the reputation value of N (R_N), the more we can trust $T_N(i)$ reported by N. Now, the answer to the above question is quite easy: we can just assign a trust value to i by combining the reported trust value $T_N(i)$ and the reputation value R_N. However, there are many ways we can combine them. So the question becomes, how do we exactly combine the reported trust and owner's reputation value. We denote the reputation value of N as R_N and the combined result as $\hat{T}(i)$. One of the easiest way to get $\hat{T}(i)$ is just to multiply R_N and $T_N(i)$. This seems reasonable at the first glance, because $\hat{T}(i)$ increases as either R_N or $T_N(i)$ increases. However, suppose N is a malicious node with very low reputation, e.g., 0.1, and N wants to send a true information item i to an end user but does not want the user to trust it, so it reports the trust value of i as 0.1. If we use this approach, the user gets

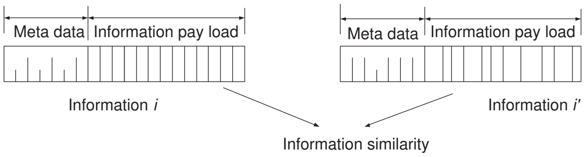

Fig. 2. Illustration of Information Similarity in a Multihop Information Flow

$$\hat{T}(i) = T_N(i) \times R_N = 0.01 \tag{1}$$

which means the user distrusts i even more than the malicious node N wanted. Therefore, this approach cannot reasonably model these kinds of situations. Instead of simply multiplying R_N and $T_N(i)$, our approach of evaluating $\hat{T}(i)$ is

$$\hat{T}(i) = (T_N(i) - 0.5) \times R_N + 0.5 \tag{2}$$

The rationale behind the above equation is, if a node's reputation is high, we tend to believe its reported trust values. However, if a node's reputation is low, no matter what trust value it reports, we tend to consider the real trust value to be 0.5, i.e., uncertain about the real trustworthiness.

We call $\hat{T}(i)$ as the *initial trust* of information item i, since it is the initial trust value we get solely based on i's provenance. If for a certain event, an end user receives only a single information item and no other related information items at all, then this is the best it can do to evaluate the trust value of this particular information item. However, if the user receives lots of information items that are trying to describe the same event, that would be another story. In the following section we will analyze these kinds of situations and adjust the initial trust by considering information items that support or conflict with each other.

B. *Information Trust Adjustment*

1) Information Similarity and Conflict: Different items about the same event in the network environment may be either supportive or conflicting. Similar information items are considered as support to each other, while conflicting information items compromise the trustworthiness of each other [10]. Now we have two questions: first, how do we determine whether two items are supportive or conflicting? and second, how to adjust the initial trust based on the similarity? The first question is not our main concern here, we propose to use a clustering algorithm to group the information items describing the same event. Therefore, each end user can have different collections of information items and each collection represents a single event. Besides, there has been lots of work done on the data similarity evaluation in the field of data mining [11], [12]. We can make use of those techniques to compare the similarity between any two information items within a collection. Illustration of information similarity is shown in Fig. 2. The information similarity comparison will be done at the payload level as the meta data will obviously be different. If the meta data is the same, then it is implied that the information item should be the same.

We assume any two information items i and i' within a collection, have a similarity of $\delta(i, i')$ ranges from -1 to 1, where -1 means completely conflicting with each other and 1 means they are exactly the same. Now we can complete the first step of answering the second question above by assigning a *similarity factor* Δ_i to information item i that belongs to a collection.

$$\Delta_i = \frac{\sum_{i,i' \in C_i, i \neq i'} \delta(i, i')}{|C_i| - 1} \tag{3}$$

where C_i is the collection that information item i belongs to.

2) Processing Path Difference: Besides information similarity and conflicts, the way that the information items are collected is also an important factor in determining the trustworthiness of the information items [10]. For example, if several independent nodes provide the same information about a particular event, such information is likely to be true. However, even if some information items are very similar to each other, but they have been processed by a large number of same nodes in their provenance path, we cannot say they are still as supportive to each other. Therefore, we need to take path correlation between any two information items into consideration as well. Illustration of path similarity is shown in Fig. 3 where the dark nodes denotes a set of common nodes between two paths.

Before we proceed to examine how to adjust trust values based on path correlation, it is important to distinguish two concepts: *Information Provenance Path* and *Information Processing Path*. We define these two terms as follows.

DEFINITION 4: **Information Provenance Path:** Information provenance path contains the entire location history of an information item and its input information items.

DEFINITION 5: **Information Processing Path:** The information processing path is the information provenance path excluding nodes which only did "pass" action on the information items.

Since those nodes which only did "pass" action do not affect the trustworthiness of the information items, what we really care about is the information processing path instead of information provenance path. In the rest of this paper, we will just use the word "path" to refer to the processing path. For two information items i and i', their path difference is:

$$PD(i, i') = \frac{max\{|P_i, P_{i'}|\} - |S\{P_i, P_{i'}\}|}{max\{|P_i, P_{i'}|\}} \tag{4}$$

where $|P_i|$ and $|P_{i'}|$ are the numbers of nodes on the paths of information items i and i' respectively and $|S\{P_i, P_{i'}\}|$ is the number of common nodes on the two paths. For information item i, we can now assign a *path difference factor* Θ_i to account for the overall path correlation with other items in the same collection.

$$\Theta_i = \frac{\sum_{i,i' \in C_i, i \neq i'} PD(i, i')}{|C_i| - 1} \tag{5}$$

3) Adjusting Initial Trust: Now we have two factors that address the similarity of information items describing the same event and the correlation of their paths. We can proceed to give a complete answer to the second question we raised earlier. The amount of adjustment to be made on the initial trust of information item i is denoted by λ_i and given as follows.

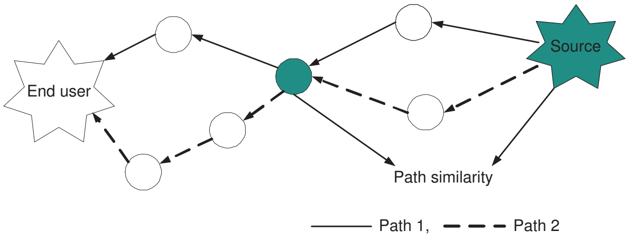

Fig. 3. Illustration of Path Similarity in a Multihop Information Flow

$$\lambda_i = \Delta_i \cdot \Theta_i \cdot e^{-\frac{1}{|C_i|}} \cdot \omega \tag{6}$$

where ω is a user-defined parameter to determine the range of adjustment we can make. The reason we introduced another term $e^{-\frac{1}{|C_i|}}$ is that the more items are in i's collection, the more we should be convinced by the similarity factor, thus the more influence this adjustment factor should have. It increases negative exponentially with the collection size because when the collection size becomes large, this effect should become smaller. We also observe that λ_i can be either positive or negative depending on whether other information items in the same collection are supporting or conflicting i. Finally, we have the *adjusted trust* $\mathbf{T}(i)$ as follows:

$$\mathbf{T}(i) = \hat{T}(i) + \lambda_i, \ 0 \leq \mathbf{T}(i) \leq 1 \tag{7}$$

C. Reputation Feedback

In addition to evaluating the trustworthiness of information, we also have the desire to update the reputation values of the network nodes. Therefore, we propose a reputation feedback approach to update the reputation of the information owner based on the adjusted trust of the information. In other words, end users give feedback to the owners after comparing information items in the same collection. In case of receiving only one information item for a certain event, the end user can not make any adjustments on the initial trust, and therefore, the reputation feedback step will not be done. However, the reputation feedback is always done when an end user receives multiple information items for a certain event and thus can obtain an adjusted trust for each information item.

As mentioned earlier in this paper, each information item has a trust value in its meta data reported by the owner. Our feedback algorithm is based on the distance between the adjusted trust and the reported trust. When the reported trust is closer to the adjusted trust than the initial trust, which means the reported trust value is more trustworthy than what we expected, we give credits to the owner's reputation. Otherwise, penalties will be given to the owner's reputation.

Note that we use different formulae to calculate R'_N depending on whether the feedback is positive or negative, i.e., credits or penalties. The reputation value increases gradually to 1 when the feedback is positive whereas it decreases linearly for a negative feedback. The idea here is that the reputation should be hard to build up, but easy to tear down [9].

Algorithm 1 Reputation Feedback

```
for all Information item $i$ has an adjusted trust $\mathbf{T}(i)$ do
    $R_N \leftarrow$ current reputation of $i$'s owner $N$
    $\rho \leftarrow$ feedback factor for updating $N$'s reputation
    $R'_N \leftarrow$ new reputation of $N$
    if $|T_N(i) - \mathbf{T}(i)| < |T_N(i) - \hat{T}(i)|$ then
        $\rho = |\lambda_i|$
        $R'_N = R_N + \rho(1 - R_N)$
    else
        $\rho = min\{|\lambda_i|, |\hat{T}(i) - \mathbf{T}(i)|\}$
        $R'_N = R_N(1 - \rho)$
    end if
end for
```

V. Performance Analysis

A. Trust Evaluation Process Analysis

First, we want to analyze how the similarity/dissimilarity of the data and paths in a collection as well as the size of the collection impact the final computed trust, i.e., the adjusted trust **T**(i). Fig. 4 shows an analysis scenario when the initial trust $\hat{T}(i)$ is a fixed value 0.7. We can see the adjusted trust varies linearly with the product of data similarity factor and path difference factor ($\Delta_i \cdot \Theta_i$), which has a range of -1 to 1. As expected, when $\Delta_i \cdot \Theta_i$ equals 0, i.e., $\lambda_i = 0$, there is no adjustment to be done, therefore the initial trust and the adjusted trust are equal. The adjusted trust value cannot exceed 1, that is why we see the lines for the adjusted trust become flat when they reach 1. Due to the term $e^{-\frac{1}{|C_i|}}$ in Eq. (6), larger collection size means larger adjustment being made, but this effect becomes smaller when the collection size continues to increase.

Secondly, we want to analyze the reputation feedback process. Fig. 5 is the plot for this analysis, which shows a scenario that the initial reputation value (R_N) is 0.6 for the information owner. In addition, we set the final trust value ($\mathbf{T}(i)$) to be 0.7. That is, when the owner reports different trust value ($T_N(i)$), λ_i changes so that the final adjusted trust will be 0.7. The blue dots are the updated reputation values (R'_N) after the feedback. We observe that when the owner reports a trust value that is far away from 0.7, it gets a negative feedback and therefore R'_N is smaller than R_N. When the reported trust value is close enough to 0.7, the owner gets a positive feedback. A counter-intuitive observation is, when the feedback is positive, the closer $T_N(i)$ is to 0.7, the smaller positive feedback we get. This is because we made the value of $\mathbf{T}(i)$ fixed in this analysis and thus λ_i changes with $T_N(i)$. When $T_N(i) = 0.7$, λ_i becomes 0, which corresponds to the case that no adjustment is to be done and hence no feedback is necessary. Therefore, the owner's reputation remains 0.6.

B. Simulation setup

We have set up a MATLAB simulation to evaluate the proposed approach. The message transmission is assumed to be packet based with each packet of size 8 bits. Messages are generated based on a source event. The event could be enemy intrusion or some natural phenomenon depending on the purpose of node deployment. The generated message packets from source nodes will be transmitted through many paths (routes) before they reach the intended destination. Therefore, the destination receives many version of the message through multiple paths. We assume a standard routing protocol exits and the message travels through 5 different paths before it reaches the destination. Two scenarios have been considered. In the first scenario, the paths are assumed to be uncorrelated, i.e., there is no common node exists between the different paths therefore $|S(P_i, P'_i)| = 0$. In the second scenario, the paths are assumed to have 2 nodes in common between them, i.e., $|S(P_i, P'_i)| = 2$. The positions of these 2 common nodes could be different among different paths. The length of the paths $max(P_i, P'_i)$ is assumed to be 6. $1,000$ packets with each packet of 8 bits binary are generated using $rand$ random number generator with changed seed value. The intermediate nodes are assumed to be either misbehavioural nodes or generating their own messages based on their own observations and also the messages they receive from the neighbours. The intermediate nodes alter the bits of the packets accordingly. This phenomenon is simulated by generating a random integer number η ranges from 0 to 8. Now every node is assumed to alter η (random number) number of bits in the message by complementing the bits. The positions of bits to be altered will again be decided randomly. The common nodes are assumed to behave consistently among all paths. The user-defined adjustment weight factor ω is set to be 1. With this simulation set-up, we obtain the following results.

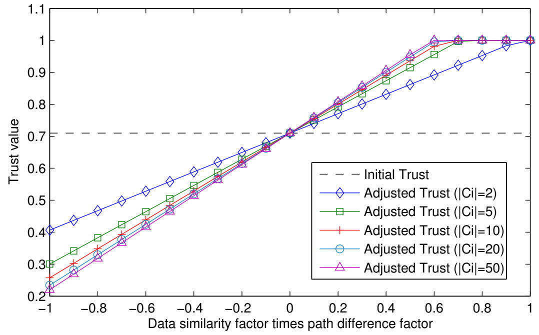

Fig. 4. Impact of Data Similarity, Path Difference and Collection Size on Information Trust

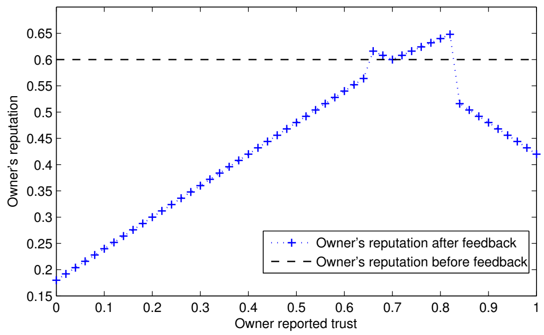

Fig. 5. Impact of Owner Reported Trust on Reputation Feedback

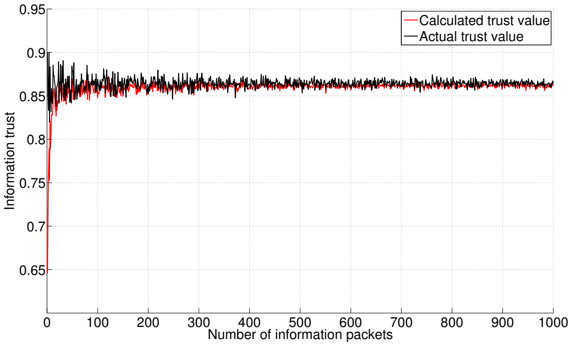

Fig. 6. Comparison of Calculated Trust Value by Simulation and Actual Trust Value

The trust value of the information is calculated based on Eq. (7). The nodes are assumed to follow the below procedures to report the trust of the information. A node compares the packet it receives from the previous node in the route path with the neighbour nodes. The ratio of average difference between these packets and the number of bits per packet (in our case it is 8) subtracted from 1 is used as the reported trust value of the packet by any intermediate nodes. Initial reputation values of all nodes are assumed to be 0.5 and the reputation values are updated using Algorithm. 1. For comparison purpose, the actual trust value of a message is calculated by comparing the original message generated at the source and the message received from each path. The differences of messages between every path with the original message is averaged. Now the ratio between the average number of different bits and the original number of bits per packet, subtracted from 1 is considered as the "*actual trust*" value. The comparison between the actual trust value and the adjusted trust value for the various number of transmitted packets when $|S(P_i, P'_i)| = 2$ is shown in Fig. 6. We can observe that as the number of packets is low, the actual trust value and the adjusted trust value have large difference between them. However, as the number of packets increases beyond 50, both the adjusted trust value and the actual trust value are closely merging. This is because the accuracy of the adjusted trust increases when we accumulate more number of evidences (informations). The reason for the fluctuations in the trust behaviours in Fig. 6 is because every node is assumed to behave randomly with every packet. However, the overall behaviour in a large scale is consistent. We observe the similar pattern even when the paths are uncorrelated i.e., $|S(P_i, P'_i)| = 0$ due to the randomness of the nodes' behaviours.

VI. Conclusion

We have proposed an information trust computation strategy based on information provenance. Our model can capture the trustworthiness of information flowing in the network as well as dynamically adjust the reputation of each network node. This scheme also allows intermediate nodes to send valuable data that are not necessarily fully trustworthy. The proposed approach is analyzed and evaluated using simulation setups. From the simulation results we can observe that the information trust obtained using the proposed method closely follows the actual information trust value. Our approach is generic and does not get influenced by mobility or any other network dynamics as long as the transmitted packet meta data contains proper provenance information. We have not considered collusion attacks and data tampering without digital signatures in this work. For testing the real efficiency of the our approach, a more detailed performance evaluation by applying our trust model in a real distributed and dynamic environment will be done in our upcoming research.

Acknowledgment

This work was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official polices, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy right notation here on.

References

[1] L. Gomez, A. Laube, and A. Sorniotti, "Trustworthiness assessment of wireless sensor data for business applications," in *AINA '09: Proceedings of the 2009 International Conference on Advanced Information Networking and Applications*, pp. 355–362, 2009.

[2] R. Shaikh, H. Jameel, S. Lee, S. Rajput, and Y. Song, "Trust management problem in distributed wireless sensor networks," in *12th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications*, pp. 411–414, 2006.

[3] T. Jiang and J. S. Baras, "Trust evaluation in anarchy: A case study on autonomous networks," in *INFOCOM-2006: 25th IEEE International Conference on Computer Communications*, pp. 1–12, April 2006.

[4] A. Boukerch, L. Xu, and K. EL-Khatib, "Trust-based security for wireless ad hoc and sensor networks," *Computer Communications*, no. 30, pp. 2413–2427, 2007.

[5] M. J. Probst and S. K. Kasera, "Statistical trust establishment in wireless sensor networks," in *Proceedings of the 13th International Conference on Parallel and Distributed Systems*, pp. 1–8, 2007.

[6] V. Lenders, E. Koukoumidis, P. Zhang, and M. Martonosi, "Location-based trust for mobile user-generated content: applications, challenges and implementations," in *Proceedings of the 9th workshop on Mobile computing systems and applications*, pp. 60–64, 2008.

[7] C. Papageorgiou, K. Birkos, T. Dagiuklas, and S. Kotsopoulos, "Dynamic trust establishment in emergency ad hoc networks," in *Proceedings of the 2009 International Conference On Communications And Mobile Computing*, pp. 26–30, 2009.

[8] B. Lagesse, M. Kumar, J. M. Paluska, and M. Wright, "Dtt: A distributed trust toolkit for pervasive systems," in *Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications*, pp. 1–8, 2009.

[9] B. Yu, S. Kallurkar, and R. Flo, "A demspter-shafer approach to provenance-aware trust assessment," in *CTS 2008: International Symposium on Collaborative Technologies and Systems*, pp. 383–390, May 2008.

[10] C. Dai, D. Lin, E. Bertino, and M. Kantarcioglu, "An approach to evaluate data trustworthiness based on data provenance," in *SDM '08: Proceedings of the 5th VLDB workshop on Secure Data Management*, pp. 82–98, Springer-Verlag, 2008.

[11] R. Weber, H.-J. Schek, and S. Blott, "A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces," in *VLDB '98: Proceedings of the 24rd International Conference on Very Large Data Bases*, pp. 194–205, 1998.

[12] Shyam Boriah, Varun Chandola, Vipin Kumar, "Similarity Measures for Categorical Data: A Comparative Evaluation," in *SIAM International Conference on Data Mining*, 2008.