Assessment of the Trustworthiness of Digital Records

Jianqiang Ma, Habtamu Abie, Torbjørn Skramstad, Mads Nygård

Assessment of the Trustworthiness of Digital Records

Jianqiang Ma[1,2], Habtamu Abie[2], Torbjørn Skramstad[1], and Mads Nygård[1]

[1] Department of Computer and Information Science,
Norwegian University of Science and Technology, Trondheim, Norway
{majian, torbjorn, mads}@idi.ntnu.no
[2] Norwegian Computing Center, Oslo, Norway
{Jianqiang.Ma, Habtamu.Abie}@nr.no

Abstract. It is easy enough to assert the trustworthiness or otherwise of a digital record, but it is far more difficult to present an objective basis for that assertion. A number of recent research efforts have focused on the trustworthiness of a digital record while paying scant attention to the record's evidential value as a measure of and a basis for the assessment of its trustworthiness. In this work, we study a model for the assessment of the trustworthiness of digital records based on their evidential values using the Dempster-Shafer (D-S) theory. The model is divided into three modules, (i) a knowledge-modelling module that models expert knowledge and consequent belief of evidence, (ii) an evidence-combination module that combines evidence from different sources in the face of uncertainty, and (iii) a trustworthiness assessment module that aggregates and integrates evidence, and assesses its trustworthiness. An example is presented to show how the model works.

1 Introduction

Due to the last century's developments in information technology, electronic documents are replacing paper documents to an ever-increasing degree. This technology enables electronic documents to be easily modified and transferred, which makes life easier for us in our digital businesses, but also makes it easier for malicious elements to compromise or tamper with them, which makes life more difficult for us. On receipt of an electronic document, one's first reaction is to question the document's trustworthiness. Current research into how to reduce the questionability of a document's trustworthiness is conducted in two areas, security and trustworthy repositories.

In the former area, in which most current research is conducted, research is concerned with the development of algorithms [17], protocols [19], and architectures [7, 8], whose purpose is to protect the electronic documents from tampering, and to ensure their trustworthiness. Even though the security methods such as digital signature and digital watermarking technology can protect digital records, they are not generally accepted in the area of digital library, as Boudrez [2] states, "*in general, the international archival community rejects the*

preservation of encrypted documents". Thus, there is still a need for a method for assessing the trustworthiness of digital records preserved in digital libraries.

In the latter area, the emphasis is on the establishment of digital repositories, the trustworthiness of which is intended to be a guarantee of the trustworthiness of the digital records stored therein [3,5]. There is, however, a need to assess the trustworthiness of the digital records themselves, since they do not reside solely in the repository at all times.

Therefore, in this work, we study the trustworthiness of the digital records, using their evidential values as a measure of trustworthiness. Specifically, we look into Evidence-Keeping Metadata (EKM) [16] related to them. The EKM are a subset of the Recordkeeping Metadata [18], but limited only to the metadata which contain the evidence to proof the trustworthiness or untrustworthiness of a digital record. Note that the digital records we studied here are the records preserved in digital library; EKM of those records are dynamically documented by the digital library system and stored in a secure place. It is assumed that the EKM are not modified. The protection of the digital library system as well as the EKM is not covered in this paper. By combining the evidential values of EKM, the evidential value of a digital record can be deduced and used to assess the trustworthiness of the record. This work is a complement to the research on both security protections and trustworthy digital repositories.

The rest of this paper is organised as follows. First, we briefly describe related work in Section 2. After illustrating the assessment approach and how to apply this approach to the assessment of the trustworthiness of digital records in Section 3 and Section 4, respectively, we present an example to show how the assessment model works in Section 5. After discussing the challenges to the assessment model which we will investigate in our future work in Section 6, we present the conclusion in Section 7.

2 Related Work

Digital trust has become an increasingly important area of research. Of special importance is the estimation of trustworthiness of information and users. Extensive surveys and overviews of trust in IT (Information Technology) can be found in [1, 9, 12, 23]. In this area, an often used methodology for trust management is the exploitation of the D-S theory of evidence [21] that defines a mathematical theory of evidence based on the belief function and plausible reasoning, which can combine separate evidence to compute the trustworthiness of an event.

There have been many attempts to apply the Dempster-Shafer (D-S) theory to the problem of assessing trustworthiness. Chen and Venkataramanan [4] applied D-S to intrusion detection in ad-hoc networks. They use D-S to combine observations on the trustworthiness of the suspected node from different nodes, and derive a number which shows the trustworthiness of the suspect node. Hu et al. [11] applied D-S to assess the trustworthiness of a digital image. They proposed a list of attributes that a digital image contains, and trained the classifier with 2000 images. Using the classifier in the experiments, their results show that

the evaluation model is stable and robust. It is evident from these attempts that the D-S theory *offers a mathematical way to combine evidence from multiple observers without the need to know a priori or conditional probabilities as in the Bayesian approach* [4]. In this study, we apply D-S to assess the trustworthiness of digital records. To the best of our knowledge, no research has been conducted on the application of D-S to the assessment of the trustworthiness of digital records using their evidential values.

3 Our Approach

In this research, we adopt the D-S theory for assessing the trustworthiness of digital records for two reasons. First, the D-S theory can combine evidence from different sources, and achieves a degree of belief based on all the available evidence into consideration [6, 21]. Second, it can handle uncertainty without requiring a priori or conditional probabilities [4, 20].

In the D-S theory there is a set of mutually exclusive and exhaustive propositions denoted by Γ, called frame of discernment. A power set 2^{Γ} contains all possible subsets of Γ, as well as the Γ itself and the null set ϕ. A mapping function from 2^{Γ} to the interval between 0 and 1 is called the basic belief assignment (or mass function), which requires that:

$$m(\phi) = 0; and \sum_{A_i \subset \Gamma} m(A_i) = 1 \tag{1}$$

The mass function $m(A)$ expresses the proportion of evidence that supports the proposition set A, but not any subsets of A. Proposition set A may contain multiple propositions due to lack of information. In this case, the mass function $m(A)$ is the source of uncertainty in the D-S theory. In this work, the evidential values of EKM are initialised by a group of experts. That is because, first, it is often difficult, if not impossible, to find an expert who has professional knowledge of the complete EKM dataset. Second, in order to have more objective results when assessing the trustworthiness of digital records, we request a group of experts, instead of a single expert, to initialise the evidential values of EKM. This is inspired by the research work in the area of instrument development [10, 15] in which a panel of experts are always used in the judgement-qualification stage so as to validate content more objectively. The basic belief assignments are used to capture experts' knowledge of EKM's evidential values, which are assigned to numeric evidential values between 0 and 1 that are converted from linguistic evidential values initialised by experts.

A belief function in the D-S theory as defined in Equation (2) is the degree of belief that the proposition set A is true. It gathers all evidence that directly supports A. If proposition set A contains a single proposition, the belief function of A equals its mass function.

$$bel(A) = \sum_{B_i \subset A} m(B_i) \tag{2}$$

A plausibility function presents the possibility that proposition set A is not negated. The plausibility function gathers all evidence that support A or do not contradict A. It is defined as:

$$pls(A) = \sum_{B_i \cap A \neq \phi} m(B_i) \tag{3}$$

$bel(A)$ and $pls(A)$ are the lower bound and upper bound of the proposition set A, respectively. They are related to each other by:

$$pls(A) = 1 - bel(\overline{A}) \tag{4}$$

In this model, the belief function is used to model the quality of EKM that provide evidence that their high-level nodes are either trustworthy or untrustworthy, while the plausibility function is used to model the quality of EKM that provide evidence that their high-level nodes may be either trustworthy or untrustworthy.

Dempster's rule of combination is used to combine the basic belief assignments from different sources. Suppose there are two sources of evidence where the basic belief assignment functions are m_1 and m_2, respectively. Then, the rule of combination is defined as:

$$m_{12}(A) = m_1(A) \oplus m_2(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \phi} m_1(B) m_2(C)} \tag{5}$$

The Dempster's rule of combination is used to combine the various assignments from different experts, and the evidence provided by the EKM, so as to assess the trustworthiness of digital records (as will be presented in Section 4). Based on the combined basic belief assignment, the corresponding belief and plausibility of the combined evidence can be obtained.

As described in Section 2, in the area of computer science, the D-S theory of evidence [21] is an often-used methodology for trust management [23]. A survey of its mathematical foundations, applications and computational analysis can be found in [14]. Since the D-S theory remains attractive because of its relative flexibility in reflecting uncertainty or lack of complete evidence and giving a convenient numerical procedure for fusing together multiple pieces of evidential data by its rule of combination [4], we use the D-S evidence theory in this paper. Despite the criticisms of the use of Dempster's rule of combination when encountering significant conflicting information [20,24], it is our considered opinion that it is suited to the assessment of the trustworthiness of digital records using evidential value as a measure of trustworthiness.

In the following section, we explain how the evidential values of EKM are used to assess the trustworthiness of a digital record in detail.

4 Assessment of the Trustworthiness of a Digital Record Using D-S Theory

In a previous work [16], we structured EKM of a digital record as a tree based on a proposed life-cycle model. The trustworthiness of the record is assessed from the

leaves to the root of the tree, where the final assessment is made. The trustworthiness assessment model is divided into three modules, (i) a knowledge-modelling module that models expert knowledge and consequent belief of evidence, (ii) an evidence-combination module that combines evidence from different sources in the face of uncertainty, and (iii) a trustworthiness assessment module that aggregates and integrates evidence, and assesses its trustworthiness.

4.1 The Knowledge-Modelling Module

In order to obtain the evidential values of EKM, a set of experts are selected to provide their knowledge about the quality of all EKM used as evidence to prove the trustworthiness or untrustworthiness of the digital record. The knowledge-modelling module models knowledge from experts to the evidential value of EKM, which will later be used to assess the trustworthiness of the digital record. Each expert assigns a tuple set (EV, H) to EKM. $EV \in$ {Extremely High (EH), Very High (VH), High (H), Medium (M), Low (L), Very Low (VL), Extremely Low (EL)} [1], presents the evidential value of EKM assigned by the expert. H is a Boolean value that stands for the "trustworthy hypothesis", if it is true, it means the expert believes that the corresponding EKM provide evidence that their higher-level nodes are trustworthy. For example, if an expert assigns (EH, true) for an EKM, it indicates that the expert thinks these EKM provide strong evidence that their higher-level nodes are trustworthy. The linguistic values can then be mapped to numeric evidential values between 0 and 1. The numeric evidential values can be presented in percentage between 0 and 100%, Table 1 shows an example of the mapping.

Table 1. Mapping linguistic evidential values to numeric evidential values.

Linguistic EV	extremely high	very high	high	medium	low	very low	extremely low	ϕ
Numeric EV	95%	80%	65%	50%	35%	20%	5%	0

After this step, the knowledge-modelling module models the experts' knowledge as a tuple set Ψ.

$$\Psi = \{\{(NEV_{11}, H_{11}), (NEV_{12}, H_{12}) \dots (NEV_{1n}, H_{1n})\} \dots \{(NEV_{m1}, H_{m1}), (NEV_{m2}, H_{m2}), (NEV_{mn}, H_{mn})\}\} \quad (6)$$

where m is the index of EKM, n is the number of experts and NEV is the numeric evidential value. Assume that H_{ij} is true, then, tuple (NEV_{ij}, H_{ij}) only means expert E_j believes that EKM_i (a piece of EKM) supports the trustworthiness hypothesis that "its higher-level node is trustworthy" to the extent NEV_{ij}. It

[1] $EV = \phi$ means that the EKM can be used to establish neither that their higher-level nodes are trustworthy nor that they are untrustworthy.

does not necessarily mean that expert E_j believes EKM_i supports the hypothesis that its higher-level node is untrustworthy to the extent $1 - NEV_{ij}$.

Note that aspects like what happen when no expert exist to provide the evidential values and the data are coming from a third party source, are not considered while it is a more realistic scenario. In addition, there is concern that some of the experts may be malicious, thus, attacks (such as bad mouthing attack, on-off attack, etc.) may be performed. In the p2p networks area, many defence mechanisms [13,22] have been proposed to detect malicious agents. Those solutions can also be adopted here to detect malicious experts. However, this will not be discussed any further in this paper.

4.2 The Evidence-Combination Module

After modelling all the experts' knowledge by the previous module, this module combines the assignments from all experts for each piece of EKM, so as to obtain the assessed evidential value of each piece of EKM.

Suppose the frame of discernment of EKM_i is $\Gamma = \{T, \overline{T}\}$, and N stands for the higher-level node of EKM_i, where T is the proposition set that $\{N$ is trustworthy$\}$, $\overline{T}$ is the proposition set that $\{N$ is untrustworthy$\}$, and U is the universal set that $\{N$ is trustworthy, N is untrustworthy$\}$. The expert E_j's assignment tuple (NEV_{ij}, H_{ij}) can then be mapped to the basic belief assignment. If $H_{ij} = \phi$, it means that EKM_i provides no evidence about the trustworthiness or untrustworthiness of N. Thus, in that case, $m_{ij}(T)$ and $m_{ij}(\overline{T})$ are equal to 0, and $m_{ij}(U)$ equals 1 (100%). In other cases, the mapping follows Equations (7) and (8) below.

$$\text{if } H_{ij} = true, \text{then } \begin{cases} m_{ij}(T) = NEV_{ij} \\ m_{ij}(\overline{T}) = 0 \\ m_{ij}(U) = 1 - NEV_{ij} \end{cases} \tag{7}$$

otherwise,

$$\text{if } H_{ij} = false, \text{then } \begin{cases} m_{ij}(T) = 0 \\ m_{ij}(\overline{T}) = NEV_{ij} \\ m_{ij}(U) = 1 - NEV_{ij} \end{cases} \tag{8}$$

As we stated at the end of Section 4.1, when H_{ij} is true, it does not mean that expert E_j believes EKM_i supports $\overline{T}$ to the extend $1 - NEV_{ij}$. Also, it is the same case when H_{ij} is false. Therefore, $1 - NEV_{ij}$ is assigned to $m_{ij}(U)$ to show that expert E_j is not certain that EKM_i supports T or $\overline{T}$ to the extent $1 - NEV_{ij}$.

By applying Dempster's rule of combination to the aggregated probabilities assigned to each piece of EKM by all experts, the evidence-combination module calculates the evidential value of each piece of EKM.

According to the mapping function presented in (7) and (8), the basic belief (i.e. $m_{ij}(T)$, $m_{ij}(\overline{T})$, $m_{ij}(U)$) assigned to each tuple is obtained. Then, using

Equation (5), the basic beliefs assigned to EKM_i by all experts are calculated as follows.

$$m_{EKM_i}(T) = m_{i1}(T) \oplus m_{i2}(T) \oplus \ldots \oplus m_{in}(T)$$
$$m_{EKM_i}(\overline{T}) = m_{i1}(\overline{T}) \oplus m_{i2}(\overline{T}) \oplus \ldots \oplus m_{in}(\overline{T})$$
$$m_{EKM_i}(U) = m_{i1}(U) \oplus m_{i2}(U) \oplus \ldots \oplus m_{in}(U)$$

Based on Equation (2), the belief functions of EKM_i are:

$$bel_{EKM_i}(T) = m_{EKM_i}(T);\ bel_{EKM_i}(\overline{T}) = m_{EKM_i}(\overline{T})$$

The belief function of EKM_i presents its evidential value. To the experts' knowledge, the quality of EKM_i that provides evidence that its higher-level node is trustworthy is $bel_{EKM_i}(T)$, while the quality of EKM_i that provides evidence that its higher-level node is untrustworthy is $bel_{EKM_i}(\overline{T})$.

4.3 The Trustworthiness Assessment Module

The trustworthiness assessment module assesses the trustworthiness of the digital record by first aggregating the evidential values of EKM to assess the trustworthiness of their corresponding components. It then integrates the trustworthiness of components to assess the record's trustworthiness during each life-cycle phase. Finally, it integrates trustworthiness during life-cycle phases to deduce the trustworthiness of the digital record.

Dempster's rule of combination is also applied here to assess the trustworthiness of the digital record. The basic beliefs assigned to the digital record are arrived at by:

$$m_{record}(T) = m_{creation}(T) \oplus m_{modification}(T) \oplus m_{migration}(T) \oplus m_{retrieval}(T)$$
$$\oplus\, m_{disposal}(T) = m_{Originator}(T) \oplus m_{Creator}(T) \oplus m_{CreationAction}(T) \oplus \ldots$$
$$\oplus\, m_{DisposalExecutor}(T) \oplus m_{DisposalAction}(T) = m_{EKM_1}(T) \oplus \ldots \oplus m_{EKM_m}(T)$$
$$m_{record}(\overline{T}) = m_{EKM_1}(\overline{T}) \oplus \ldots \oplus m_{EKM_m}(\overline{T})$$
$$m_{record}(U) = m_{EKM_1}(U) \oplus \ldots \oplus m_{EKM_m}(U)$$

Accordingly, the belief and plausibility of the trustworthiness of the digital record are calculated respectively as follows:

$$bel_{record}(T) = m_{record}(T); \quad pls_{record}(T) = 1 - bel_{record}(\overline{T}) = 1 - m_{record}(\overline{T})$$

The belief function states that we can believe that the digital record is trustworthy to the extent that $bel_{record}(T)$. The plausibility function states that the digital record may be trustworthy to the extent that $pls_{record}(T)$. The trustworthiness of the digital record is a value within the interval from $bel_{record}(T)$ to $pls_{record}(T)$. However, to be conservative, we say that the trustworthiness of the digital record is $bel_{record}(T)$.

5 An Example of the Assessment of the Trustworthiness

In this section, we present an example showing how the assessment model works based on a proposed record's life-cycle model [16] with elaborated EKM required for the assessment. Here, we only describe the trustworthiness assessment during the creation phase as an example of the trustworthiness assessment of a digital record, since the approach is basically the same for the other life-cycle phases.

Suppose three experts E_1, E_2, and E_3 share their knowledge by assigning evidential values to EKM, as shown in Table 2. The numeric evidential values are given right after the linguistic evidential values and expressed as percentage for clarity.

Table 2. Assigned evidential values of EKM during creation.

	EKM	**Exp.1**	**Exp.2**	**Exp.3**
Originator	Name	(EH (95%)), true)	(EH (95%), false)	(ϕ (0), ϕ)
	Affiliation	(VH (80%), true)	(L (35%), false)	(L (35%), false)
	Compose Time	(ϕ (0), ϕ)	(H (65%), false)	(H (65%), true)
Creator	Name	(H (65%), true)	(H (65%), false)	(VH (80%), true)
	Affiliation	(VH (80%), true)	(M (50%), false)	(L (35%), false)
Creation	Record's Name	(VH (80%), true)	(VH (80%), false)	(VH (80%), true)
	Time	(L (35%), true)	(H (65%), false)	(VH (80%), false)
	Environment	(H (65%), true)	(VH (80%), true)	(H (65%), false)
	Format	(EH (95%), true)	(EL (5%), true)	(M (50%), false)
	Source	(H (65%), true)	(H (65%), true)	(VL (20%), false)
	Reason & Purpose	(VH (80%), false)	(VH (80%), true)	(H (65%), false)

By mapping linguistic evidential values to numeric evidential values, the experts' knowledge during the creation phase is modelled as follows:

$$\Psi_{creation} = \{\{(95\%, true), (95\%, false), (0, \phi)\} \ldots \\ \{(80\%, false), (80\%, true), (65\%, false)\}\}$$

Then, using Equations (7) and (8), the basic beliefs of EKM during the creation phase can be assigned. As an example, the basic belief assigned to "name of originator" is given below:

$$m_{ONname_1}(T) = 0.95; \quad m_{ONname_1}(\overline{T}) = 0; \quad m_{ONname_1}(U) = 0.05$$
$$m_{ONname_2}(T) = 0; \quad m_{ONname_2}(\overline{T}) = 0.95; \quad m_{ONname_2}(U) = 0.05$$
$$m_{ONname_3}(T) = 0; \quad m_{ONname_3}(\overline{T}) = 0; \quad m_{ONname_3}(U) = 1$$

The evidence-combination module then combines the experts' knowledge for each piece of EKM using Equation (5) as follows:

$$m_{ONname}(T) = [m_{ONname_1}(T) \oplus m_{ONname_2}(T)] \oplus m_{ONname_3}(T) \approx 0.4872$$
$$m_{ONname}(\overline{T}) \approx 0.4872; \quad m_{ONname}U \approx 0.0256$$

Since both proposition sets T and $\overline{T}$ only contain a single proposition, its belief function equals its mass function, as presented in Equation (2). Thus, $bel_{OName}(T) = m_{OName}(T) = 0.4872$, $bel_{OName}(\overline{T}) = m_{OName}(\overline{T}) = 0.4872$.

The belief function states that, based on the three experts' knowledge, the reliability score of the originator's name as evidence of the truth of the hypothesis "the originator is trustworthy" is 48.72%. It also states the reliability score of the originator's name as evidence of the hypothesis "the originator is untrustworthy" is 48.72%. This occurs due to the experts' conflicting knowledge and opinions. We address this issue in Section 5.1.

The combined results of all EKM during the creation phase are presented in Table 3.

Table 3. Combined results of experts' knowledge on EKM during creation phase.

	EKM	$m(T)$	$m(\overline{T})$	$m(U)$
Originator	Name	0.4872	0.4872	0.0256
	Affiliation	0.6282	0.2147	0.1571
	Compose Time	0.3939	0.3939	0.2122
Creator	Name	0.8230	0.1150	0.062
	Affiliation	0.5652	0.2717	0.1631
Creation	Record's Name	0.8276	0.1379	0.0345
	Time	0.0363	0.8962	0.0675
	Environment	0.8230	0.1150	0.062
	Format	0.9094	0.0453	0.0453
	Source	0.8514	0.0297	0.1189
	Reason & Purpose	0.2188	0.7266	0.0546

The trustworthiness assessment module aggregates the evidence of attributes to their parent components to assess the trustworthiness of those components, as shown below.

$$\begin{aligned} m_{Originator}(T) = 0.6779; \quad m_{Originator}(\overline{T}) = 0.3197; \quad m_{Originator}(U) = 0.0024 \\ m_{Creator}(T) = 0.8918; \quad m_{Creator}(\overline{T}) = 0.0940; \quad m_{Creator}(U) = 0.0142 \\ m_{Creation}(T) = 0.9848; \quad m_{Creation}(\overline{T}) = 0.0152; \quad m_{Creation}(U) = 0 \end{aligned}$$

Based on the aggregated results, it integrates the trustworthiness of components into the parent level, where the trustworthiness of the digital record during the creation phase is obtained, as shown below.

$$\begin{aligned} & m_{creation}(T) = 0.9991;\ m_{creation}(\overline{T}) = 0.0009;\ m_{creation}(U) = 0 \\ & bel_{creation}(T) = 0.9991;\ pls_{creation}(T) = 1 - bel_{creation}(\overline{T}) = 0.9991 \end{aligned} \tag{9}$$

Equation (9) states that, based on the experts' knowledge of the EKM during the creation phase, the trustworthiness score of the digital record is 99.91% after its creation.

Similarly, the trustworthiness of the digital record for the other phases of the life-cycle can be calculated. Finally, the trustworthiness of the digital record can be assessed by integrating its trustworthiness in all its life-cycle phases.

Below, we highlight three cases to present how the trustworthiness assessment of the experts' knowledge works.

5.1 Case One

As in the above example, Expert 1 and 2 agree that the originator's name has extremely high evidential value, which means it is a strong evidence for the trustworthiness or untrustworthiness of the component Originator. However, the knowledge of each individual expert concerning the trustworthiness hypotheses is in conflict with that of the other(s). Combining their knowledge, we arrived at $bel_{Exp_1 \& Exp_2}(T) = 0.4872$, $bel_{Exp_1 \& Exp_2}(\overline{T}) = 0.4872$, and $bel_{Exp_1 \& Exp_2}(U) = 0.0256$. As the belief functions show, due to the contradictory nature of the knowledge of the experts, although it is strong evidence in the experts' opinions, it can not be used to support the assertion of either the trustworthiness or untrustworthiness of the record in any way that has any high evidential value.

5.2 Case Two

To continue with the calculation of the trustworthiness of originator's name, due to the lack of information, Expert 3 believes that the originator's name can prove nothing about the trustworthiness or untrustworthiness of the component. Thus, he/she assigned basic belief with full uncertainty ($m_{Exp_3}(U) = 1$). When combining the knowledge of Expert 3 with the knowledge of other experts, assignments from Expert 3 have no impact on the combined result.

5.3 Case Three

About the EKM "affiliation of originator", Expert 1 believes that it supports the claim that the originator is trustworthy, while other experts have opposing opinions. If their knowledge is combined based on the majority-vote approach, the results will suggest that the originator is untrustworthy. However, from the perspective of Expert 1, the affiliation of the originator has very high evidential value, which means the evidence it presents is strong. The other experts regard the affiliation as having low evidential value, which means although it suggests that the originator is not trustworthy, the evidence is not strong enough. Thus, the combined results using D-S theory suggest that affiliation provides evidence that the originator's trustworthiness score is 62.82%, which acknowledges Expert 1's knowledge.

6 Discussion and Future Work

There have been many researches on the area of digital trust, however, they either focus on developing secure algorithms [17], protocols [19], and architectures [7, 8] to protect digital records, or pay attention to the establishment of

trustworthy repositories [3,5], which are intended to be a guarantee of the digital records stored therein. To our knowledge, no research has been conducted into the calculation of the trustworthiness of digital records using evidential value as a measure of trustworthiness. In addition, not much research has been conducted into the value of the metadata around digital records as evidence of the trustworthiness of these records. Therefore, we look into the EKM of digital records. By using the D-S theory of evidence, we developed a model for the assessment of the trustworthiness of digital records. Our model demonstrates that the incremental improvement of experts' knowledge in the area of evidential value, and the adoption of a rigorous formal approach, make possible the objective assessment of the trustworthiness of digital records.

There still remain a number of challenges to be met, including (1) the impact on the assessment from the temporal aspect is a challenge which needs further research, (2) as one of the criticisms on the D-S theory, the way of handling conflicts between EKM in the assessment model needs further studies, (3) since EKM may have different importance to the assessment result, weighting difference within the model needs further investigation, and (4) some of EKM can be interrelated in the model, such as name and affiliation of an operator, thus, how to combine dependent EKM needs further studies.

7 Conclusion

In this paper, we have developed and described a model for the assessment of the trustworthiness of digital records using Dempster-Shafer (D-S) theory. It uses the records' evidential values as a measure of trustworthiness. This model consists of three modules, (i) a knowledge-modelling module, which models the experts' knowledge related to a digital record, (ii) an evidence-combination module, which combines experts' knowledge of evidential values of Evidence-Keeping Metadata (EKM), and (iii) a trustworthiness assessment module, which assesses the trustworthiness of the digital record by aggregating and integrating evidence of EKM. We have presented an example with three cases to show how this model works. We have also identified challenges to the assessment model, which we will investigate in our future work.

Our results show that by incrementally improving experts' knowledge about evidential values and applying a rigorous formal approach, the trustworthiness of digital records can be assessed objectively.

As mentioned in the previous section, in our future work, we will continue to investigate the temporal, conflict, weighting and dependency aspects of the trustworthiness assessment of digital records.

References

1. Blaze, M., Feigenbaum, J., Ioannidis, J., Keromytis, A.D.: The role of trust management in distributed systems security. In: Secure Internet Programming: Issues in Distributed and Mobile Object Systems. LNCS, vol. 1603, pp. 183–210 (1999)

2. Boudrez, F.: Digital singatures and electronic records. Archival Science 7(2), 179–193 (2007)
3. Center for Research Libraries: Trustworthy repositories audit & certification: Criteria and checklist (2008), http://www.crl.edu/PDF/trac.pdf
4. Chen, T.M., Venkataramanan, V.: Dempster-Shafer theory for intrusion detection in ad hoc networks. IEEE Internet Computing 9(6), 35–41 (2005)
5. Consultative Committee for Space Data Systems: Reference model for an open archival information system (OAIS) (2002)
6. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
7. Gladney, H.M.: Trustworthy 100-year digital objects: Evidence after every witness is dead. ACM Transaction on Information System (TOIS) 22(3), 406–436 (2004)
8. Gladney, H.M., Lorie, R.A.: Trustworthy 100-year digital objects: durable encoding for when it's too late to ask. ACM TOIS 23(3), 299–324 (2005)
9. Grandison, T., Sloman, M.: A survey of trust in internet applications. IEEE Communications Surveys and Tutorials 3(4) (2000)
10. Grant, J.S., Davis, L.L.: Selection and use of content experts in instrument development. Research in Nursing & Health 20(3), 269–274 (1997)
11. Hu, D., Wang, L., Zhou, Y., Zhou, Y., Jiang, X., Ma, L.: D-S evidence theory based digital image trustworthiness evaluation model. In: MINES 2009 (2009)
12. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems 43(2), 618–644 (2007)
13. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in p2p networks. In: WWW 2003. pp. 640–651 (2003)
14. Kohlas, J., Monney, P.A.: Theory of evidence - a survey of its mathematical foundations, applications and computational aspects. Mathematical Methods of Operations Research 39(1), 35–68 (1994)
15. Lynn, M.: Determination and quantification of content validity. Nursing Research 35(6), 382 (1986)
16. Ma, J., Abie, H., Skramstad, T., Nygård, M.: Development and validation of requirements for evidential value for assessing trustworthiness of digital records over time. Journal of Information (2011), (to appear)
17. Menezes, A., Oorschot, P., Vanstone, S. (eds.): Handbook of applied cryptography (1996)
18. National Archives of Australia: Australian government recordkeeping metadata standard. Tech. rep. (2008)
19. Rescorla, E.: SSL and TLS: Designing and building secure systems. Addison-Wesley Professional (2000)
20. Sentz, K., Ferson, S.: Combination of evidence in dempster-shafer theory. Tech. rep. (2002)
21. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
22. Sun, Y.L., Han, Z., Yu, W., Liu, K.: Attacks on trust evaluation in distributed networks. In: CISS 2006. pp. 1461–1466 (2006)
23. Trcek, D.: A formal apparatus for modeling trust in computing environments. Mathematical and Computer Modelling 49(1-2), 226–233 (2009)
24. Zadeh, L.A.: A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination. AI Magazine 7(2), 85–90 (1986)