Structure and Evidence in Identity Cases

Emma Sloan
Department of Computer Science
Brown University
Providence, RI
emma_sloan2@brown.edu

Marguerite McDaniel, William Nick
Dept. of Computer Sci.
North Carolina A&T State University
Greensboro, NC
{mamcdan2, wmnick}@aggies.ncat.edu

James Mayes
Department of Political Sci, & Criminal Justice
North Carolina A&T State University
Greensboro, NC
jpmayse@ncat.edu

Albert Esterline
Dept. of Computer Sci.
North Carolina A&T State University
Greensboro, NC
esterlin@ncat.edu

***Abstract*— We present a framework for agent identity with a focus on structured cases and numerical levels of evidence and their handling. Our framework focuses on id-situations, where a person is judged to be the agent in some scenario, particularly a crime scene. An id-situation has a constellation of associated situations (providing what we call an id-case) to produce the objects used in it. Our idea of a situation is modeled on Barwise and Perry's situation theory. We represent our situations using semantic web notation because, for one thing, the semantic web supports a structure of partial information. From a numeric standpoint, we create a justification-based mass function from each id-situation and then combine the multiple functions we get from having multiple id-situations using Dempster-Shafer theory. The id-situations give a measure of similarity between each suspect involved and the as yet unknown criminal, which we adapt to get a mass function, with a frame of discernment that can be approximated as the list of suspects. We then refine our frame of discernment using constraints based on objects shared between the id-situation and its corresponding supporting situations.**

Keywords—identity; Dempster-Shafer theory; semantic web

I. INTRODUCTION

Identity is a complex, multifaceted problem with aspects ranging from the philosophical question of how we define ourselves to the practical, legal question of how we identify criminals. We consider only some aspects of identity, specifically how we specify information about persons, rather than the deeper philosophical issues involved. To notate that information, we enhance the framework for identity introduced by Dominguez et al. [1]. Our framework examines how we put together persons' actions and interactions with others, particularly in a criminal justice centric scenario. The fundamental question is how we support identity judgments.

Computational forensics [2, 3] is relevant to the work reported here in a general way in that it applies computational methods to forensic science, where forensic science is described as "the methodological correct application of a broad spectrum of scientific disciplines to answer questions significant to the legal system" [2]. Although our goal is a computational framework, most of the more fundamental issues we address arise whether or not computational resources are used.

Our framework focuses on situations, and we represent situations using semantic-web standards (RDF, etc.), we have developed OWL ontologies for this research, and we use these standards in drawing inferences. We are not, however, concerned at this point with issues of how identity in general may be represented on the web (cf., e.g., `owl:sameAs`) as is, for example, [4]. Referential opacity, for example, is a notion that we are saving for future work. What distinguishes our work from general work on identity and the web is that we are concerned with the identity of agents both in cyber and physical environments. We are particularly concerned with provenance of information and how a case fits together to support an identity judgment.

The SuperIdentity project is the current state-of-the-art in frameworks for agent identity [5]. The project is interdisciplinary, considering the psychological and forensic aspects of identity, a focus on the relation between identity and the internet. To the authors' knowledge, the SuperIdentity project is the only other project that provides a general framework for identity in which evidence for identity of any kind is combined. The SuperIdentity framework starts with some known information or *element of identity*, such as a username or email address, and *transforms* that element into others, such as by looking up an email address to find the associated username or phone number. If this information is represented as a directed graph, the transformations form the edges, and each element is a node. The elements are grouped by type, such as all usernames, into *characteristics*, multisets of elements of the same type. The set of all characteristics is a person's *superidentity*, an organized version of all known information on the individual.

The SuperIdentity project also provides a visual representation of how much any factor of identity can be used to access another, called the reachability matrix [6]. Each column in the reachability matrix is a known element of identi-

THE AUTHORS ACKNOWLEDGE SUPPORT FROM THE NSF (GRANT NO. 1460864) AND THE ARO (CONTRACT NO. W911NF-15-1-0524)..

ty or type of information, while each row is an unknown the user is trying to find. Each unknown is given a rating of ease and accuracy of accessibility when starting from the element in each column. For example, you can find an email address from a user's social media profile that is visible to friends with high accuracy and medium effort.

Any superidentity has a measure of certainty for each individual element and the superidentity as a whole. An element's confidence is taken from its provenance; the confidence measures of every element and transformation that came before the element in question are multiplied to determine its confidence. Once characteristics are created, those element confidence measures can be modified based on how similar an element is to others within its characteristic. The quality of the entire superidentity is determined from its number of characteristics, lack of conflict, and reinforcement, or amount of connection between the sets of nodes in the graph interpretation.

Our framework will cover all aspects of the SuperIdentity framework but from a situation-focused perspective rather than starting with some known part of the identity. This different perspective lets us handle more physical elements of identity, such as biometrics, while the SuperIdentity project focuses primarily on online forms of identity. We also approach measures of confidence in both the individual elements and the superidentity as a whole differently.

The remainder of this paper is organized as follows. The next section covers background, including situation theory, in which we formulate our cases involving identity, and data semantics, which provides an improvement on some aspects of situation theory. The background also includes semantic web resources (such as RDF), which we use to encode our information, query it, and reason about. Finally, background includes Dempster-Shafer theory, which we use to reason about evidence and uncertainty. Section 3 presents our running example, and the next section explains how we apply Dempster-Shafer theory, in both combining and refining evidence. Section 5 discusses partial objects postulated in data semantics that grow, possibly together, as information grows, thus providing a dynamic model for identity. Section 6 discusses our implementation, concentrating on retrieving information for an id-case and calculating the level of evidence for various alternatives. Section 7 concludes and suggests future work.

II. BACKGROUND

A. Situation Theory

Our framework organizes information around situations as described by Barwise's situation theory. (See [7], the reissue of the 1983 text that launched situation theory, and [8], Devlin's systematization of situation theory.) A real situation is much like the layman's definition: some happening or state in the world, which may contain an unbounded number of pieces of information. In contrast, an abstract situation is a well-defined subset of that state, made up of a finite number of infons, formalized pieces of information. Infons are notated as $\langle\langle R, a_1 \ldots a_n, l, t, i \rangle\rangle$. *R* stands for a relation, which connects a number of objects a_1 through a_n. That relation between the objects occurs at a given time, *t*, and location, *l*. *i* is the polarity, which is 0 if things are not thus related and 1 if they are. For example, the statement "the book is on the table" expresses an infon with relation *on*, objects *book* and *table*, and a positive polarity, if given a time and location. Infons can be parameterized, with one or more objects left unspecified. Situations *support* infons if the infon is true in the given situation.

A situation can imply the existence of another situation, which forms a *constraint*. Constraints can come from nature, language, or societal conventions. Any sort of implicitly understood signal forms a constraint. For example, that an elevator beeping means you have reached your floor is a constraint; the situation of a beeping elevator implies the existence of a situation with that elevator reaching a given floor. Another example constraint is a verbal description of a situation, such as saying "I saw the cat climb the tree." The person stating the fact is in one situation, an *utterance situation*, which is tied with a constraint to the enclosed *described situation*, which is the cat climbing the tree.

We are especially concerned with situations where someone in authority judges the identity of some agent (e.g., the culprit in a crime). We call such a judgment an *id-action* and the situation an *id-situation*, which is an utterance situation with, e.g., the crime scene the corresponding described situation. As we explain below, there are typically several situations that together support the id-action; we call these *supporting situations*. The id-situation, the described situation, and the supporting situations together make up what we call an *id-case*. We consider any id-action to be the assertion of an identity statement even when it contains non-linguistic elements. For example, pointing at someone and saying "He did it" identifies one person in two ways: as the target of pointing and as the object denoted by the description (with an understanding of what "it" denotes in this context). Identity is an equivalence relation, and a collection of asserted identities defines equivalence classes. All the denoting devices in each class denote the same individual, and it is these equivalence classes that correspond in our approach to superidentities.

B. Data Semantics

While situation semantics models specific happenings in the real world, Landman's [9] data semantics models facts that interpret the statements uttered by language users in a conversation. These facts are atomic propositions, describing relations between objects similarly to infons. Landman considers a collection of these facts to be a representation of the world, which is shared in conversation. We can allow possible facts in our representation of the world. Some possible facts are not true; they are simply facts that are not compatible with others. Information states are sets of compatible facts and are partially ordered, defining a lattice. (Regarding this partial order, propositions are partially or-

dered by a relation of information containment, and this ordering on propositions induces a partial order $\subseteq$ on information states such that, where $\mathbf{s_1}$ and $\mathbf{s_2}$ are information states, $\mathbf{s_1} \subseteq \mathbf{s_2}$ iff $\mathbf{s_1}$ contains all the information that $\mathbf{s_1}$ contains and possibly more.) An information state grows as the conversation or investigation proceeds. How an information state grows is constrained by conditional statements, which are similar to the constraints in situation theory. Landman's account of information growth and conditionals is particularly appealing since it provides a model for how the information from supporting situations is integrated into the id-situation. We have worked out a way to accommodate situation theory within data semantics, by enhancing each atomic statement with an argument denoting the situation where it holds.

C. Semantic Web Resources

We notate situations using semantic web resources. Statements are given as syntactic "triples" *subject property object*, as specified by the Resource Description Framework (RDF) [10]. The subject and object of the triple must be Universal Resource Identifiers (URIs), literals, or bnodes, which act as placeholders for some existent objects. URIs are unique across the web, and can be written more generally as URIrefs, each of which has a prefix and a local identifier. For example, the triple **`id:314 rdf:type foaf:Person`** has as subject an employee URIref, and the property says that the subject is of the type specified by the object, here meaning that the URIref represents a person. The RDF resource descriptions are saved in a triple store.

RDF Schema (RDFS) lets us define new RDF vocabularies [11]. We can specify information about a given property in RDFS, such as its range and domain. RDFS also groups resources into classes and allows for subclass relations. RDFS can apply set-theoretic relations such as union and intersection through subclass or subproperty relations. The Web Ontology Language (OWL) is an extension of RDFS that provides further classifications of properties and classes. (We use OWL 2 [12].)

We define a number of classes and properties to notate situation theory with these semantic web resources. We define subclasses of class **`:Infon`** for specific types of infon, such as **`:AnalystMatchingInfon`** for where an analyst is matching crime scene evidence to biometric information from our suspects. The type of infon specifies the relation, and each of the objects involved in the infon are connected to the infon with properties. This lets us capture n-ary relations even though RDF properties are binary.

Triple stores can be queried using SPARQL, which is similar to SQL but works with triples. SPARQL uses a SELECT clause to extract multiple values that fulfill specifications given in a `WHERE` clause. We use the Jena Semantic Web framework [13] to write SPARQL queries in Java.

Semantic web resources are useful because they allow a freedom of expression matching the vast possibilities of identity. The idea that anyone can say anything about any topic is a motto for the semantic web [14]. We use this flexibility to ensure that any facet of identity, whether on or offline, is describable in our framework.

D. Dempster-Shafer Theory

We use Dempster-Shafer theory to provide a numerical measure of confidence in our identities. It distributes and combines justification-based evidence that assigns masses to sets of elements, with the total mass summing to 1.0 [15]. Any set of elements, including singleton sets with only one element, with some non-zero mass assigned is called a focal element. Evidence can also be given to the set of all possible elements, which is the frame of discernment. This mass assignment differs from likelihood-based evidence, as used in probability theory, primarily in that masses in probability theory are only ever assigned to one individual at a time, essentially a singleton. Justification-based evidence allows for a preservation of uncertainty, represented by the assignment of mass to the entire frame of discernment.

Frames of discernment can be analyzed to create new, more detailed frames in refinements [16]. The analysis provides a more detailed set of information but not necessarily a larger or smaller one. For example, one refinement of the frame {A, B, C} is {A, B, C^D, C^¬D}, which provides two more specific categories without changing the total information from all focal elements.

Dempster-Shafer theory also has functions to get a lower or upper bound on a set's likelihood, respectively, belief and plausibility. The *belief* associated with a set is determined by adding the masses of all of its subsets. For a frame of discernment Θ, a mass function m, and any subset θ of Θ, $Bel(\theta) = \sum_{\theta^* \subseteq \theta} m(\theta^*)$, which sum is in [0,1] as the sum of the masses of all focal elements is 1.0. The plausibility of a set is the sum of the masses of all sets that overlap with it or, for a frame Θ, a mass function m, and any subset θ of Θ, $Plaus(\theta) = \sum_{\theta^* \text{ s.t. } \theta^* \cap \theta \neq \emptyset} m(\theta^*)$.

We use Dempster-Shafer theory to combine evidence. It provides a number of rules to combine mass functions while maintaining uncertainty. For specifics of combination rules see Section IV, Work with Dempster-Shafer Theory.

III. Running Example

We consider a legal case in which a theft has occurred immediately following a party, providing a list of possible suspects in the form of a guest list. Evidence from the crime scene reveals a group photograph from a security camera with one guest with their hand on the door to where the valuables were kept and a fingerprint from that same door. This case can be described as a constellation of situations, centering around two separate id-situations for the two pieces of evidence: the fingerprint and the snapshot.

Situation s_1 is the id-situation for the fingerprint, in which an analyst compares fingerprints on file from the par-

tygoers to the forensic one found at the crime scene. See Figure 1. To compare the fingerprints, each one, whether from the crime scene or the police department, is taken and handled in its own situation. Specifically, in all the situations numbered s_{3a}-s_{3d}, a suspect has their fingerprint taken by a police officer, and in situation s_4, the criminal touches the doorknob, placing the forensic fingerprint. The crime scene investigation team then lifts the fingerprint from the doorknob for usage in this case in situation s_5.

Similarly, the id-situation for the security camera image, s_2, is supported by its own constellation of situations. See Figure 2. Police take a mugshot of each suspect in situations s_{6a}-s_{6d}. Those mugshots are then compared with the security camera image in s_2. The security camera records the group in situation s_7, which acts as an utterance situation, describing the actual occurrence of a group standing near the door and someone touching the doorknob while at the party, s_8. Situation s_4, in which the fingerprint is left by touching the doorknob, is a part of situation s_8, the larger scene of what occurred at that moment at the party.

We consider a situation to provide our frame of discernment, specifically the situation in which the criminal is identified, parameterized to handle the different suspects. That situation would be an utterance situation, as in a legal scenario, a prosecutor would be asserting the criminal's identity. The frame can also be approximated as the group of suspects themselves, allowing us to put a mass on an individual or group of individuals rather than a situation.

In s_1 and s_2, additional fingerprint and facial photographs are taken for matching. For such searches to be done, search warrants are usually required. To obtain such warrants, law enforcement must demonstrate that the "situation" satisfies the "probable cause" requirement of the Fourth Amendment of the US Constitution. There are examples, however, where probable cause is evident and search warrants are not needed. For example, a security pack in a bag of cash stolen from a bank will "explode," marking the cash in red. Being covered in red dye would imply the person's involvement in in the robbery and would be a constraint providing sufficient probable cause for additional searches.

IV. Work with Dempster-Shafer theory

We now detail how Dempster-Shafer theory is used to combine and modify mass functions to get numerical indications of how the evidence supports alternative identity judgments. We consider each id-situation and its constellation of associated situations to create its own mass function and provide evidence for the likelihood that each suspect was the criminal. Our scenario (Section III) provides a numerical similarity measure between each suspect and the crime scene for each piece of evidence, which is turned into a mass function and normalized to sum to 1.0. Those mass functions would then be combined according to Dempster-Shafer theory, as discussed in the following subsection, Combining Mass Functions. We then refine the frames of discernment and therefore modify the mass functions to handle the metadata for each similarity measure, which is discussed in the second subsection, Refinement and Weighting.

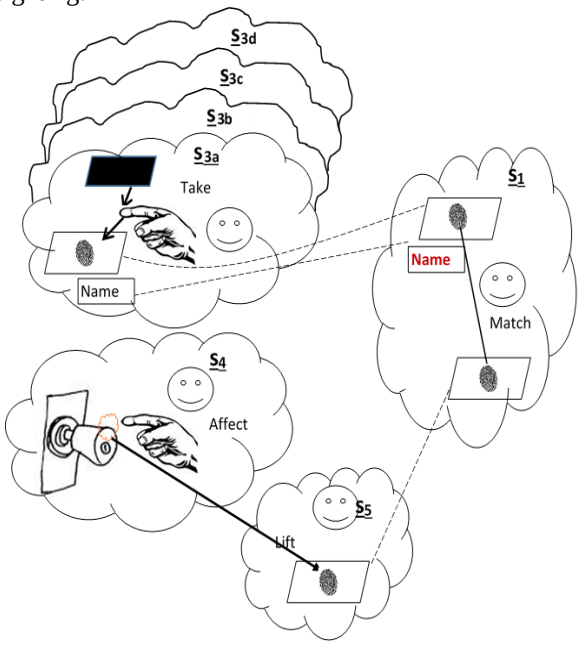

Figure 1: The fingerprint id-case

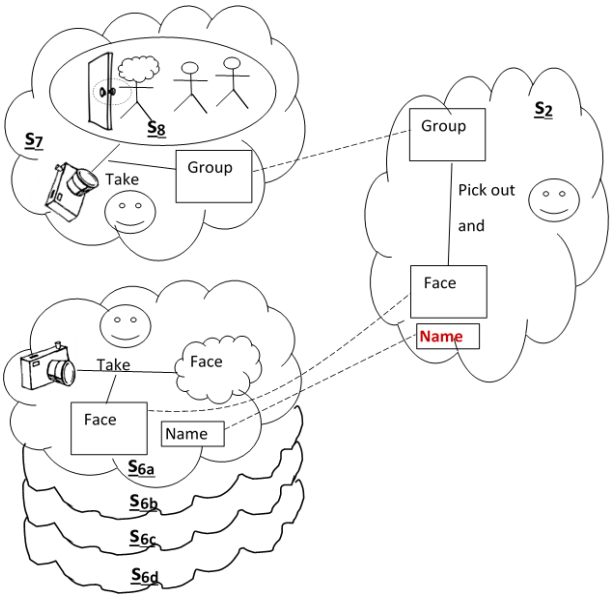

Figure 2: The mugshot id-case

A. Combining Mass Functions

We use Dempster's rule for combining mass functions, which divides conflict between the different mass functions evenly among these elements and does not set it as uncertainty [17]. Dempster's rule calculates a measure of conflict $K = \sum_{B \cap C \neq \emptyset} m_1(B) m_2(C)$ for each pair of mass functions m_1 and m_2 and all focal elements B and C. The combined mass function according to Dempster's rule is $m_{12}(A) =$

$\frac{\sum_{B\cap C=A} m_1(B)m_2(C)}{1-K}$, where 1-*K* (with *K* as defined above) is for normalizing. We chose this combination rule because conflict among different pieces of evidence is likely not indicating that the suspect is unknown when evidence comes from dissimilar sources. This rule works reasonably well for the running example in this paper, but in other settings, other combination rules may be more useful.

Dempster's rule directly contrasts with a similar combination rule, Yager's rule [18]. Its combined mass function would be $m_{12}(A) = \sum_{B\cap C=A} m_1(B)m_2(C)$ for all focal elements except the frame of discernment, which has *K* (representing conflict) added. Since it applies all conflict directly to uncertainty, it would be fitting, for example, in a scenario where we believed some or all of the evidence to have been planted to blame an innocent for the crime.

In our running example, we assume a limited list of possible suspects, which gives us the same suspects and therefore the same frame of discernment in both the evidence from the fingerprint and from the mugshot. However, in other cases, a piece of evidence could lead to new suspects not suggested by previous evidence. Therefore, it would be appropriate to consider each piece of evidence as its own frame of discernment. One existing combination rule, Zhang's center combination rule [19], takes in evidence from two separate frames of discernment. However, Zhang's rule takes in a very specific structure of known information and would not be appropriate to apply to our framework without some adaptation of either the framework or the rule because we have no obvious relation between our different sources of evidence and therefore between our potential frames of discernment.

B. Refinement and Weighting

Once we have determined our frame of discernment, we refine it. Our refinements come from constraints, similar to what was done by Lalmas and van Rijsbergen [20]. However, while their constraints are linguistic, based on the nature of their work with documents, our constraints are conventional and come from the objects used to form evidence, which are shared between situations. By convention, if an object is used in evidence, it must have been properly collected and handled. In a legal scenario, this is called the chain of custody and must be complete for evidence to be admissible in a court, unless there is a witness to identify the evidence [21]. In our framework, the collection, copying, and handling of the evidence form situations, refining the frame of discernment. The people involved in each step may or may not be reliable, so the masses are recalculated, lowering for unreliable handling, with the removed mass going to uncertainty.

Our refinements were created to follow axioms that cover both the mathematical aspects of Dempster-Shafer theory, such as that all masses sum to 1.0, as well as the legal and practical aspects of the framework, such as that the mass on a particular suspect should not increase without crime scene evidence suggesting guilt. We plan to express the axioms in informal logic to support inference.

Some parts of the chain of custody affect the reliability of the entire piece of evidence, rather than the mass on any particular suspect. Any crime scene evidence that was planted or in some other way unreliable would set the entire piece of evidence in question, even if the information from the suspects being compared to it was perfectly reasonable. From a legal perspective, this would be considered to be a break in protocol. This could be accounted for by using a Dempster-Shafer combination rule that allows each piece of evidence to be given a weight, such as the mixing rule [FK02]. These weights would add to 1.0 in total and allow us to classify some evidence as more valuable than other pieces. For mass functions m_1 through m_n and weight functions w_1 through w_n and focal element *A*, the mixing combination rule is $m(A) = \frac{1}{n}\sum_{i=1}^{n} w_i m_i(A)$. In future work, we plan to incorporate elements of this combination rule to create a more nuanced one. We will use argument schemes, patterned on the work by [22], to create a combination rule that precisely fits our framework.

V. PEGS

A particularly attractive feature of data semantics is how it handles information growth, and, indeed, it provides a dynamic approach to identity. We have a mix of known and unknown information about our culprit, and the known information grows as we discover more. We represent this as a partial object, specifically a *peg*. According to Landman [9], pegs have properties attached to them by the information state, an approximation of the world based on language and the atomic formulas being shared by language users. Pegs are incompletely described, with some known information and some given unknowns [9]. When those unknowns are discovered, the peg grows, and it can merge with other pegs if they are found to be the same. Each piece of evidence is likely to reveal properties of our culprit, so we continually grow our peg. The world as described by pegs is semantic, coming from shared information, which contrasts with the physical world mapped by infons.

We can describe our culprit using the sematic web to notate pegs. A URI, defined perhaps by a police department, allows the culprit to be uniquely determined but also possibly combined with other pegs. The semantic web does not assume that a URI is a unique identifier of the piece of information; this assumption is called the non-unique name assumption as discussed in [14]. So the URI hung on a particular peg does not have to be the only URI hung on said peg. In fact, two pegs can be easily merged when represented in this way using the **`owl:sameAs`** property. In idactions, we combine multiple pegs, setting up an equivalence relation between the peg for the culprit and the peg representing the suspect determined to be the criminal. We can also specify that two pegs are in fact separate by using the **`owl:differentFrom`** property. Inverse functional

properties (of class **owl:InverseFunctionalProperty**) play a special role for identity: if P is an inverse functional property and we have **a P x** and **b P x**, we can infer **a owl:sameAs b.**

VI. Implementation

Our scenario provides a distance measure for facial images and for fingerprints. When, for example, the fingerprint from the crime scene is compared with that of a suspect, a vector of the values of certain features of each fingerprint is used. These feature vectors are identical in type: they are the same length, and values in corresponding positions are for the same feature. All values are in the interval [0,1]. We calculate the Manhattan distance between the two feature vectors and normalize by dividing by the sum of the maxima in each dimension so that the value is in [0,1], where identical documents would have a distance of 0. (See [23] for details and justifications.) To create a mass function from these distance measures, we first extract the distance measures from the triple store using SPARQL and manipulate them in Java. The following SPARQL is a fragment of the code used for that extraction.

```
SELECT ?num ?distance
 WHERE
 ?infon sitterms:distanceMeasure ?distance .
 ?infon sitterms:fpRecorded ?fp .
 ?rec biom:hasFpImage ?fp .
 ?num recterms:hasRecord ?rec .
```

This code retrieves a suspect's identification number and an associated distance measure by determining the infon in which the distance measure is recorded, finding the associated fingerprint, tracking the criminal record that includes said fingerprint, and finding the identification number of the suspect with that record. With the help of our colleagues in the CASIS Center at North Carolina A&T State University, we came up with realistic sets of distances for fingerprint and facial-image matching. Comparison of interclass and intraclass distances suggested a threshold of 0.65; a fingerprint or facial image farther from the target than this is discarded as a non-match. (Again, see [23] for details.) We obtained extremely similar results from both uniformly generated distance measures between 0 and 1 and more realistic distance sets, so we use the random measures from here on.

These distance measures are then transformed into masses using a transformed sigmoid function $y = 1/e^{-8(-x+0.65)}$. Compared to usual sigmoid functions, the argument is scaled by -8 so that values fall in the range [0,1] and the value decreases with increasing x (so that greater distance results in less mass). We shift right by 0.65 for a more gradual threshold at 0.65, but actually just discard matches with distances over 0.65. We must still then normalize to get a sum of masses of 1.0. The results of this initial mass function are shown in Table 1.

The frame of discernment used to calculate those masses is then refined by following the chain of custody for the fingerprints. We as yet do not include any information from the chain of custody of the forensic fingerprint or the one initially located on the doorknob. Instead, we follow the chain of custody of each suspect's fingerprints. The fingerprints were compared by a forensic analyst and taken by a forensic professional, both of whom have some measure of reliability between 0 and 1.0 given to them by the police department, stored in our triple store as a property of the officer. We multiply the masses by these reliability measures and move the removed mass to the frame of discernment, as uncertainty. (Since the mass is being moved to uncertainty rather than removed altogether, no renormalization is necessary.) See Table 1. The reliability measures used in Table 1 and throughout this example were randomly generated in [0.7,1.0].

Once the mass function for the fingerprint data is finalized, a similar process occurs on the group photograph. The distance measures are also retrieved, put into a sigmoid

Suspect ID	Fingerprint Distance	Initial Mass	Analyst reliability	FP taker reliability	Final Mass
201	0.430	0.373	0.800	0.980	0.292
202	0.660	0.000	0.800	0.860	0.000
203	0.490	0.342	0.910	0.810	0.252
204	0.570	0.286	0.800	0.860	0.197

Table 1: Mass values obtained from fingerprint evidence, along each step of their determination. The final mass for the entire frame of discernment is 0.228, which must be included for the final mass to sum to 1.0.

Fingerprint Evidence			
Suspect	Mass	Belief	Plaus
201	0.292	0.292	0.551
202	0.000	0.000	0.259
203	0.252	0.252	0.511
204	0.197	0.197	0.456
All	0.259	1	1
Photographic Evidence			
Suspect	Mass	Belief	Plaus
201	0.503	0.503	0.678
202	0	0	0.175
203	0.322	0.322	0.497
204	0	0	0.175
All	0.175	1	1
Combined Evidence			
Suspect	Mass	Belief	Plaus
201	0.532	0.532	0.605
202	0	0	0.073
203	0.338	0.338	0.411
204	0.056	0.056	0.129
All	0.073	1	1

Table 2: Mass, belief, and plausibility (abbreviated to plaus) measures for the three mass functions created by our two id-situations and their combination using Dempster's rule

function to get initial masses, and then modified based on refinements from the chain of custody of the mugshots. The two mass functions from the two pieces of evidence are combined using Dempster's rule, as shown in Table 2.

We also consider the case where less precise information on the suspects gives us non-singleton mass assignments. Working with the same scenario and data for fingerprints, we assume that we have a photograph where the culprit's face is not clearly visible and all we have to go on is that the culprit is blond. A human would need to assign masses here based on inspection of the suspects rather than a more precise distance metric. As suspects 201 and 204 (and no others) are blond, the mass function for the photograph assigns a mas of 0.8 to {201, 204} and 0.2 to the entire frame of discernment. Table 3 shows the combined mass, belief, and plausibility measures for a fingerprint with mass assigned as above and this mass function for photographs. In this scenario, we assume that masses are assigned by a human inspecting the photo. A major distinction between photos and fingerprints is that a photo depicts a person while a fingerprint depicts only the friction ridges on a person's finger. A photo, but not a fingerprint, may thus provide information distinguishing a set of persons on a common property.

Focal element	Mass	Belief	Plausibility
{201,204}	0.26	0.872	0.937
{201}	0.366	0.366	0.69
{203}	0.063	0.063	0.128
{204}	0.247	0.247	0.571
All	0.065	1	1

Table 3: Combined mass, belief, and plausibility measures for a fingerprint with mass assigned as above and a photograph that transmits very little data and only differentiates between a set of possible suspects (201 and 204) and less likely suspects.

From the final mass functions both here and in Table 2, we can draw a number of conclusions. The simplest one would be to set up a threshold of guilt, such as 0.5 belief, over which any focal elements would be suspected of being guilty. However, we can also determine which factors were contributing and note how much each step along the process mattered. For example, this series of mass functions could easily reveal corruption if a suspect's mass was much higher before reliability measures were taken into account. We are working on a web interface to make our software available to students in the Criminal Justice Program at North Carolina A&T State University to provide experience with crime-scene evidence, and we hope to make this framework easily accessible to legal and forensic professionals and provide those conclusions readily.

VII. Conclusion

We present an expanded framework of identity, with a focus on numerical levels of evidence and their handling. Our framework focuses on id-situations, in which a person is determined the agent in some scenario, particularly a crime. These id-situations each have a constellation of associated situations to produce the objects used in the id-situation. Our idea of a situation is modeled on Barwise and Perry's situation theory. We represent our situations using semantic web notation because, for one thing, the semantic web supports a structure of partial information.

From a numeric standpoint, we create a justification-based mass function from each id-situation and then combine the multiple functions we get from having multiple id-situations using Dempster-Shafer theory. The id-situations give us a measure of similarity between each suspect involved and the as yet unknown criminal, which we adapt to get a mass function, with a frame of discernment that can be approximated as the list of suspects. We then refine our frame of discernment using constraints based on shared objects between the id-situation and its corresponding supporting situations.

Turning to future work, we plan to expand the framework in a number of ways. The work with data semantics and pegs discussed here will ideally be explored further and combined more completely with the usage of Dempster-Shafer theory. We also plan to refine our usage of Dempster-Shafer theory in a number of ways, such as creating our own combination rule specific to our framework. It is possible that our framework should be adapted somewhat to work well with Dempster-Shafer theory as well as the reverse, particularly in finding non-singleton focal elements that make sense with the legal aspect of the framework.

We have considered a number of applications once the framework is somewhat more fleshed out. The next steps in terms of application would be continued work with the criminal justice arena by getting specific data and scenarios from actual past cases, rather than randomly generating our data. Our framework could quite easily be applied to digital forensics, as the semantic web notation we use was designed for handling online information.

As mentioned, to the authors' knowledge, the SuperIdentity project is the only other project that provides a general framework for identity, but we approach identity differently as we focus on situations that make up a case for identity and equivalence classes defined by identity statements rather than sets of "characteristics" making up a superidentity. Yet we cover or plan to cover all aspects of their framework that we mentioned, especially measures of identity quality and confidence, which we handle with Dempster-Shafer theory.

To some extent, our project and the SuperIdentity project have different aims. A superidentity is all known information on an individual, organized into characteristics, each of which consists of identity elements of the same type. These elements are connected by transformations that overall form a directed graph exhibiting provenance. We, in contrast, consider the analogue of a superidentity to be an equivalence class defined by identity judgments. Each such judgment is supported by a case in the legal sense, and a case is analyzed as a constellation of situations. Our concern

with the details of a case contrasts with the SuperIdentity project's broad-stroke construction of characteristics as multisets of elements and superidentities as sets of characteristics. Each approach enjoys advantages when the required level of detail is appropriate for it. While the more detailed approach has the advantage of actually building cases, the less detailed approach would allow more individuals and "elements of identity" to be considered and could avoid constitutional issues regarding access to information to which the more detailed approach is subject.

ACKNOWLEDGMENT

This research is based upon work supported by the U.S. Government including the National Science Foundation. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

[1] Y. Dominguez, W. Nick, and A. Esterline, "Situations, identity, and the Semantic Web," in *2016 IEEE Int. Multi-Disciplinary Conf. on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2016, pp. 109-115.

[2] K. Franke and S. N. Srihari, "Computational forensics: An overview," in *International Workshop on Computational Forensics*, 2008, pp. 1-10.

[3] S. N. Srihari. (2010, December 2010) Beyond CSI: The Rise of Computational Forensics. *IEEE Spectrum*. 38-43.

[4] H. Halpin, P. J. Hayes, J. P. McCusker, D. L. McGuinness, and H. S. Thompson, "When owl:sameAs isn't the same: An Analysis of Identity in Linked Data," in *Int. Semantic Web Conf.*, 2010, pp. 305-320.

[5] D. Hodges, S. Creese, and M. Goldsmith, "A model for identity in the cyber and natural universes," in *2012 European Intelligence and Security Informatics Conf. (EISIC)*, 2012, pp. 115-122.

[6] S. Creese, M. Goldsmith, J. R. Nurse, and E. Phillips, "A data-reachability model for elucidating privacy and security risks related to the use of online social networks," in *11th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications*, 2012, pp. 1124-1131.

[7] J. Barwise and J. Perry, *Situations and attitudes*, Reissue ed. Stanford, CA: CSLI Pubs., 1999.

[8] K. Devlin, *Logic and information*: Cambridge U. Pr., 1995.

[9] F. Landman, *Towards a theory of information: The status of partial objects in semantics*. Riverton NJ: Foris Publications USA, 1986.

[10] P. J. Hayes and P. F. Patel-Schneider. (2014). *RDF 1.1 Semantics, W3C Recommendation*. Available: http://www.w3.org/TR/2014/REC-rdf11-mt-20140225/

[11] D. Brickley and R. Guha. (2014). *RDF Schema 1.1, W3C Recommendation* Available: http://www.w3.org/TR/2014/REC-rdf-schema-20140225/

[12] W3C Owl Working Group. (2009). *OWL 2 Web Ontology Language*. Available: http://www.w3.org/TR/2012/REC-owl2-overview-20121211/

[13] The Apache Software Foundation. (2013, Mar. 20, 2014). *Apache jena*. Available: http://jena.apache.org

[14] J. Hendler and D. Allemang, *Semantic Web for the Working Ontologist*: Morgan Kaufmann, Burlington, MA, 2008.

[15] J. Y. Halpern, *Reasoning about uncertainty*: MIT Press, 2005.

[16] G. Shafer, *A mathematical theory of evidence*: Princeton Univ. Pr., 1976.

[17] G. Shafer, "Probability Judgement in Artificial Intelligence," in *First Conf. on Uncertainty in Artificial Intelligence (UAI1985)*, New York, 1986.

[18] R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information sciences*, vol. 41, pp. 93-137, 1987.

[19] L. Zhang, "Representation, independence, and combination of evidence in the Dempster-Shafer theory," in *Advances in the Dempster-Shafer theory of evidence*, R. R. Yager, J. Kacprzyk, and M. Fedrizzi, Eds., Wiley, 1994, pp. 51-69.

[20] M. Lalmas and C. Van Rijsbergen, "Situation theory and Dempster-Shafer's theory of evidence for information retrieval," in *Incompleteness and Uncertainty in Information Systems*, V. S. Alagar, S. Bergler, and F. Q. Dong, Eds., Springer, 1994, pp. 102-116.

[21] Findlaw for Legal Professionals. (2001). *Summary of the Rules of Evidence: The Essential Tools for Survival in the Courtroom*. Available: http://corporate.findlaw.com/litigation-disputes/summary-of-the-rules-of-evidence.html.

[22] Y. Tang, N. Oren, S. Parsons, and K. Sycara, "Dempster-shafer argument schemes," *Proc. of ArgMAS*, 2013.

[23] A. Alford, K. Bryant, T. Abagez, G. V. Dozier, J. C. Kelly, J. Shelton, *et al.*, "Genetic and evolutionary methods for biometric feature reduction," *International Journal of Biometrics*, vol. 4, pp. 220-245, 2012.