Evidential Reasoning for Radiological Detection

Final Report

Neil Walkinshaw
Department of Informatics
The University of Leicester

Funded by an AWE NUSeC pilot project grant.

Abstract

Detecting radioactivity is inherently uncertain. In a typical passive detection scenario (e.g. scanning cargo coming off a ship), a large number of factors need to be taken into account, all of which can be subject to varying levels of uncertainty. For example, sensor count rates can vary, cargo can contain benign sources of radiation (e.g. bananas or tiles) that could mask other proscribed materials, the shipping manifest might not be clear, etc.

Ultimately, this uncertainty tends to be handled on a subjective basis. Individual operators take what are often ad-hoc decisions. This means that there can be a large degree of variation in terms of how operators handle detection events such as alarms.

In this project we attempt to apply a statistical multi-faceted decision support technique called Evidential Reasoning (ER) to provide a more systematic means by which to address this uncertainty. ER represents a decision problem as a hierarchy of factors, ranging from the top overarching question, down to the low-level atomic factors. It enables an operator (or indeed a sensor depending on the context) to supply an opinion on each of the lowest-level sub-factors, and automatically amalgamates this information to provide a high-level outcome. Importantly, this outcome makes any doubt or uncertainty explicit and quantifiable.

1 Introduction

The detection of radiation can be challenging, especially when in an uncontrolled environment. Even in its simple form (without considering the wider circumstances), detection is inherently a statistical activity [TL15]. The decision as to whether radiation exceeds a given threshold is invariably the result of taking a reading over a period of time, and producing an average, which is necessarily associated with a degree of error. Any reading is thus invariably associated with a degree of uncertainty.

This uncertainty is however invariably compounded when one considers the broader settings in which detection takes place. Readings can be confounded by the masking presence of other benign radioactive materials that could have a masking effect. They can be hidden by shielding materials. Reliability of readings can degrade at longer distances from the source, and this degradation can vary depending on the type of sensor used. The distance can itself be a source of uncertainty in a dynamic environment where the radiation source (or sensors) are in motion.

These (and other) factors make the ultimate decision of whether a radiation source has been detected or not a highly complex one. When all of these sources of uncertainty are factored in on an ad-hoc basis, there is the danger of producing large numbers of false-positive or even false-negative decisions. It is reasonable to argue that this is a particular danger in non-proliferation settings, where an adversary would seek to exploit this uncertainty (e.g. with the extensive use of masking and shielding materials).

This project is concerned with the application of a support technique for such complex decisions, known as Evidential Reasoning (ER) [YX02]. ER provides a framework within which complex multi-factor assessments can be assessed in simple, probabilistic terms. ER is particularly useful because it enables the assessor to, for specific facets of the decision, explicitly factor in any doubt or absence of information.

To facilitate our research, we have applied and refined our EVIRE Evidential Reasoning tool. This is the product of research conducted over the past four years. So far we have successfully applied it to the complex problems of reason about the safety of safety-critical software systems [NWK14, NWKdlV15] and, through an on-going DSTL-funded project, to reasoning about the trustworthiness of (semi-) automated decisions taken by UAVs.

2 Evidential Reasoning

Evidential Reasoning (ER) [YX02] provides a means by which to assimilate multiple assessments of individual facets of the evidence into a single, coherent macroscopic assessment. For this section, we use the task of software safety assessment as a basis for illustrating some the core ER concepts.

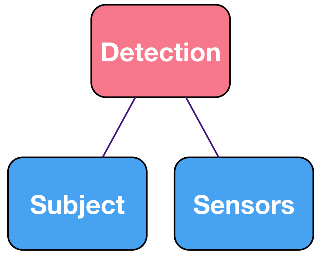


Figure 1: “Tiers” of evidence, where higher-level attributes are influenced by lower-level attribtes.

ER considers tiers of ‘attributes’ by which the evidence is to be assessed. For example, as a higher level attribute one might consider the quality of the personnel/team to be a factor, which could in turn be decomposed into sub-attributes such as competence or experience.

Traditionally, these levels of factors are represented as a hierarchy. However, it is also possible for two higher-level attributes to share the same lower-level attribute. In other words, the decision structure can be a Directed Acyclic Graph and does not have to be tree-shaped.

Formally, each attribute can be subdivided into a set E of lower-level sub-attributes $e_1, \square, e_n$. Each sub-attribute e_i can also be given a weight w_i representing the relative importance, such that $\sum_{(i=1)}^{n}(w_i = 1)$ (by default the weight is evenly distributed across all attributes as $\frac{1}{n}$).

A human expert assesses each of the lowest-level attributes by providing their assessment of the “quality” of the attribute. This is provided in terms of a Likert-scale consisting of g grades $H =< H_1, \ldots, H_g >$ (i.e. if $g = 5$, H_1 and H_5 might correspond to “very poor” and “excellent” respectively). Instead of providing just a single value on this scale (implying that they are 100% certain of their assessment – e.g. that the experience of the team is ‘excellent’), the assessment can be provided as a distribution. This distribution is referred to as ‘Belief Function’ – a term that will be used throughout the paper. This belief function enables the assessors to capture any uncertainty that they might have about their assessment. For example, an assessor could indicate that they have a confidence of 50% that the team’s level of experience is “excellent” and 50% that the level of experience is “good”.

Formally, the expert´s confidence that a particular attribute e_i achieves a grade H_n is denoted $\beta_(n, i)$. For a given attribute, $\sum_(i = 1)^n(\beta_(n, i) \leq 1)$. Thus, an expert´s complete assessment of attribute e_i (encompassing all possible grades) can be expressed as the distribution:

$$S(e_i) = \{(H_n, \beta_(n, i)), n = 1, \ldots, g\}$$

A key feature of ER is that, alongside uncertainty, it is also possible to capture complete ignorance on the part of the assessor. The beliefs in $\beta_(n, i)$ do not have to sum up to 1 for a given attribute (as would be expected with conventional Bayesian probabilities). The sum of beliefs $\sum_(i = 1)^n \beta_(n, i)$ can be interpreted as their overall confidence of the assessment, where a sum of 1 amounts to total confidence, and a sum of 0 amounts to total ignorance (no confidence).

Given a hierarchy of attributes, where the lowest-level attributes are associated with distributions corresponding to the assessments as presented above, ER provides an algorithm by which to assimilate them. Distributions of 'beliefs' are propagated up from lower-level nodes to higher-level nodes, and are combined with the distributions from their sibling nodes to produce a representative macroscopic assessment.

Crucially, this process of propagation from basic attributes to an aggregated result y obeys certain desirable axioms that ensure the following [YX02]:

1. y must not be assessed to a grade H_n if none of its basic attributes is assessed to a grade H_n.
2. y should be precisely assessed to a grade H_n if all of its basic attributes are precisely assessed to a grade H_n.
3. If all of the basic attributes are completely assessed to a subset of evaluation grades then y should be assessed to the same subset of grades.
4. If an assessment of any basic attribute is incomplete, then the assessment for y should also be incomplete to a certain degree.

3 ER-Driven Decision Support for Radiological Detection

In this section we show how ER can be applied to the task of radiological detection to facilitate decision making. Detection is problematic because it is (1) multi-faceted, there are many sources of data to consider, and (2) it is fraught with uncertainty. Many aspects of a detection decision are necessarily based upon data that is either partial (even absent altogether) or subject to error. The use of evidential reasoning can help to standardise these complex decision making procedures. It also provides a means by which to make any uncertainty explicit in the final decision.

A concrete scenario that was discussed during the course of the project was the potential to apply ER to detection decisions in the context of shipping ports. In this context, vehicles arriving in the UK are subject to scans, often by scanners that span individual lanes. Detection decisions are challenging in this environment for several reasons:

- Shipping containers or trucks can contain a large variety of **benign sources of radiation**. For example bananas contain relatively high doses of potassium, and tiles (especially terracotta tiles or Italian ceramics) can contain high levels of various radioactive nuclides.

- Containers or trucks provide **a lot of scope for shielding**. This can prevent the detection of photons, which are used as markers for the identification of isotopes.

- Benign sources of radiation are a key source of **'nuisance alarms'**, which can lead to the expensive and time-consuming task of opening up a crate and carrying out a detailed inspection.

- Sensors can give **varying baseline readings**, depending on where they are situated. For example, if the ground under one lane contains slightly radioactive buried materials (as can often be the case at ports), then its readings will need to be interpreted in a somewhat different light to other sensor readings.

- **Manifest logs are not necessarily sufficiently detailed or trustworthy**. Shipping containers can contain mixtures of materials from various sources, and these contents and sources are rarely recorded in a detailed manner.

- The threshold for deciding whether or not to check a truck / container **can vary according to the current threat-level**.

The end-result is that any detection outcome has to be carefully interpreted. Alarms triggered in certain circumstances will carry more credence than others. This variability places the bulk of the responsibility on the subjective intuition of the operator. This leaves open the risk that a relatively inexperienced operator can take the wrong actions. At best this means spending unnecessary time and money reacting to 'nuisance' alarms. At worst it means that proscribed radiological materials pass are not apprehended.

In this scenario, the use of ER could provide a basis for 'operationalising' detection decisions. Data would be presented to the operator in such a way that the various sources of data are interpreted and aggregated in a consistent way. There is less of an onus on the operator to rely upon their intuition. It is also easier to standardise the decision process.

We use this case study to set out how ER *could* be applied. Naturally, this is merely a proof-of-concept; the actual decision structure and the manner in which data is provided to the system would need to be established with input from domain experts. Nevertheless, it does serve to illustrate the general process of applying Evidential reasoning. Of starting from the basic decision structure, and figuring out potential ways by which to formulate belief functions.

3.1 Border Control Case Study

We are operating a portal-monitor in a single lane for lorries disembarking from ships at a port. There is negligible background radiation. We have access to manifest data. The portal monitor is of the standard type described by Tsoulfanidis and Landsberger [TL15]; it includes scintillation detectors to detect gamma radiation, as well as ^{3}He neutron detectors to detect gammas / neutrons emitted by enriched uranium or plutonium.

The port is extremely busy, a large load of trucks are disembarking the vessel, scanners on other lanes are out of operation, which means that there is pressure to process the trucks as rapidly as possible. For every truck, the operator has to assess whether or not it could possess proscribed radioactive materials. If there is an alarm, she also has to decide whether or not to act upon it. This amounts to opening up the truck and inspecting its contents – causing delays to the disembarkation process and diverting resources.

3.2 Application of ER

In this scenario, ER would be applied by presenting the data for each truck to the scanner operator as an aggregated decision structure. The operator would be able to visualise the aggregated data on a computer monitor (or a mobile device such as a tablet).

The decision structure would be static, agreed in advance (according to operational guidelines provided by Border Force). The supply and encoding of data as belief functions would also be encoded in advance, though certain aspects would need ot be calibrated for each portal (e.g. to account for portal-specific background radiation levels).

3.2.1 Decision Structure

We start by setting out a decision structure. The specifics of this structure depend on the specific policies and priorities set out by Border Force. For this case study we assume that the decision should explicitly factor in all of the considerations listed above.

A possible decision structure (and the one that we will adopt for our case study) is shown in Figure 2. The top node is the overall assessment; how trustworthy is a particular source (truck or crate in our case). This is broken down into the evidence we are receiving from the scanners, and the evidence we have about the source itself (via the shipping manifest). There are two types of scanners on our portal – an array of scintillation detectors and ^{3}He detectors. For each type of detector we examine the reading, and the reliability of that reading.

3.2.2 From Data to Belief Functions

Once the belief structure has been finalised, a key challenge is to provide input in the form of belief functions. To make any ER-driven tool practical, these would have to be either (a)

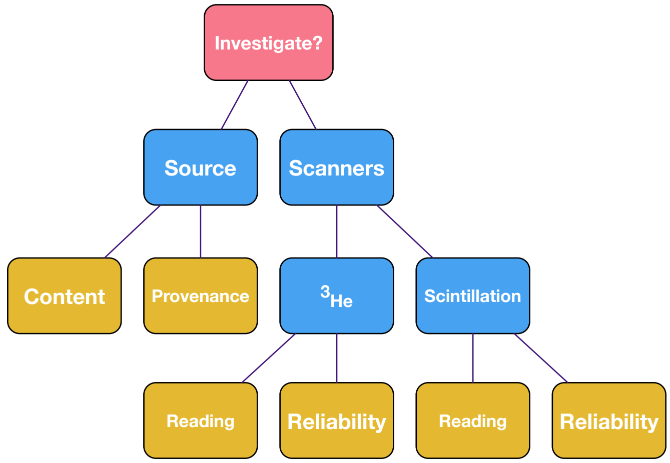


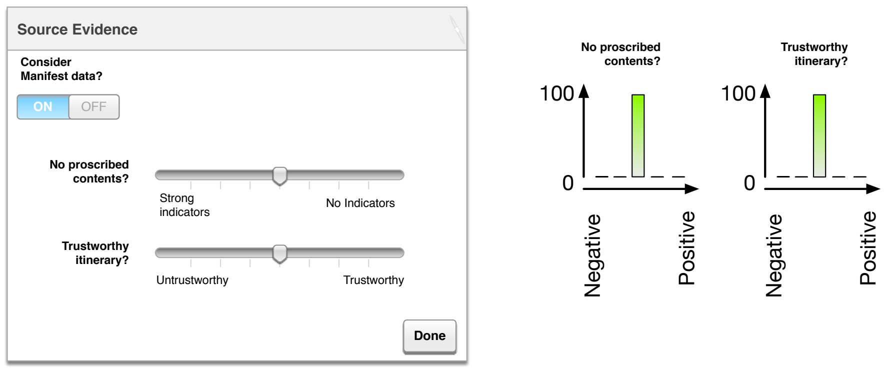
Figure 2: Detection decision

automatically generated from data or (b) provided by a human operator through a simple interface. For each of the leaf-nodes in the decision tree in Figure 2 we consider how this might be achieved.

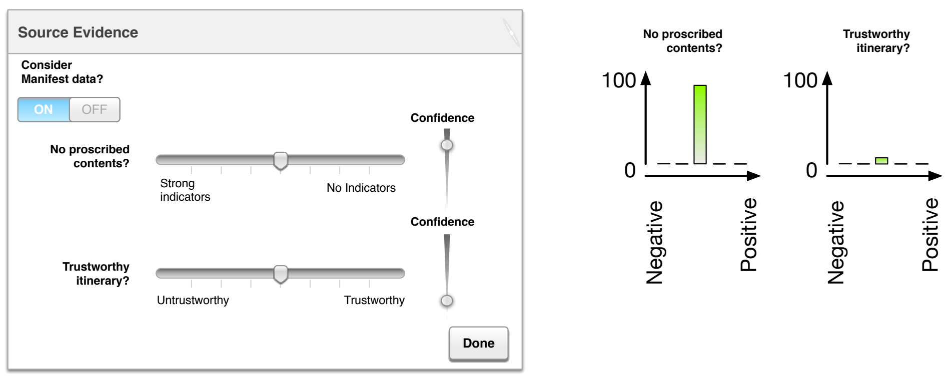
Source content and provenance From the manifest data, how confident can we be that the contents do not contain proscribed radioactive material? To provide the data, the operator would ideally be presented with a simple, intuitive GUI. Two possible GUIs are shown in Figure 3.

The first GUI is the simpler version. It provides two slider-bars, where the operator can (based purely on their interpretation of the manifest text) gauge their confidence in the fact that (a) there are no proscribed contents in the truck / container, and (b) that the itinerary does not raise any suspicions. If the slider is moved to the right, then this signals a confidence that there are no proscribed items or that the itinerary is fine respectively. As shown to the right, entries via this GUI can be straightforwardly converted into a belief function. In this case, the slider is moved half way along, which corresponds to a score of three out of five, which is simply given a belief mass of 100%.

In this case, there is no scope for accounting for any *doubt*. If the text in the manifest is too terse to fully convey what is in the container, the operator might not as confident of their assessment as they would be with a full and detailed description. To enable any doubt to be taken into account explicitly, the second GUI in Figure 3(b) also contains a slider that enables the operator to gauge their confidence in their assessment. In this case, the belief function is constructed in a similar fashion to the previous GUI, but the confidence slider is used to



(a) Simple GUI



(b) GUI with ability to indicate confidence for each factor.

Figure 3: Possible GUI designs for interpreting shipping manifest data.

attenuate the belief mass for the given score.

Scanner reading and reliability From the data obtained from the portal scanner, how confident can we be that the truck / container contains no proscribed materials? For each type of scanner, two forms of belief function are provided: Reading and Reliability. “Reading” conveys an interpretation of the scanner output (e.g. 0= no radiation, 5=lots of radiation). “Reliability” conveys the confidence in the reliability of reading – i.e. the absence of any screening materials or masking radiation.

The ”Reading” data can be obtained automatically from the scanner. How this is done depends on the type of scanner. However, if the scanner simply returns a count-rate, one could create a mapping from ranges of count rates to the scale 1–5. The certainty of a score could also be calculated according to the error associated with the count rate (if obtainable).

The reliability could be entered by the operator, having considered the various factors that

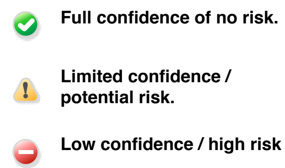


Figure 4: Possible output recoding

can confound a reading – the speed of the vehicle, the knowledge of any possible masking radiation sources or shielding materials etc. This confidence could be entered by a similar GUI to the ones shown in Figure 3.

3.2.3 Output Interpretation

By default, the output of the ER algorithm is itself a belief function, which is computationally aggregated from the evidence fed in via the leaf nodes. Each belief function is accompanied by an 'uncertainty' score (ranging from 0 to 1, where 1 is total uncertainty). In our tool, this uncertainty is also shown as a colour gradient from red to white, where red represents complete uncertainty and white represents complete certainty.

In an applied context such as our cargo-scanning scenario, interpreting a decision tree could be subject to interpretation. This can be addressed by providing alternative visualisations. For example, a Belief function could be reduced to a single "traffic-light" (see Figure 4). How this is done depends on the measurement context and Border Force protocols, but one example would be to have the following rules:

- Green: $Uncertainty < 0.1$ and $> 90\%$ of the belief mass in the belief function is spread within levels 4 and 5.
- Amber: $0.1 < Uncertainty < 0.5$ and $> 80\%$ of the belief mass in the belief function is spread within levels 4 and 5.
- Red: Otherwise (neither of the above rules is satisfied).

3.3 Vignettes

To illustrate the approach in action, using the above decision structure, we consider three trucks driving through the portal, each of which would trigger one of the three levels of alert, described previously.

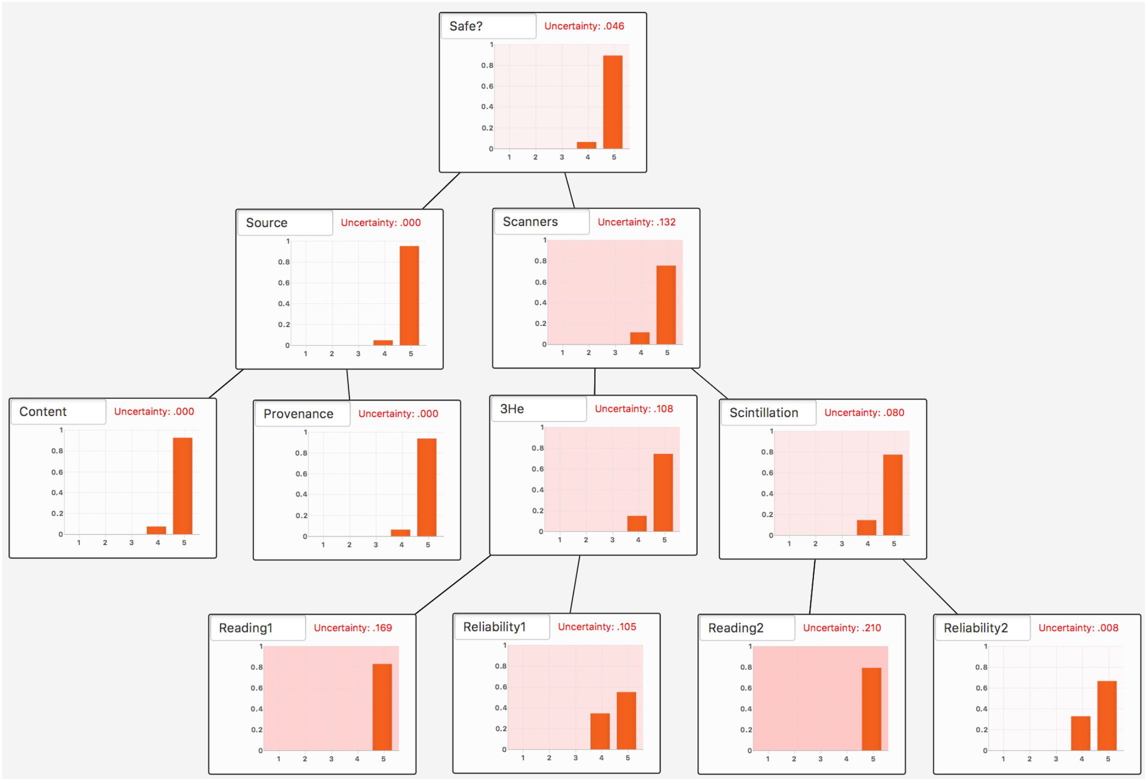


Figure 5: Sample decision tree computed for Truck A.

Green / no alert and trustworthy - Truck A Truck A does not set off any sensors. The readings are clear, there was plenty of time to capture the readings. The shipping manifest is detailed and stamped. Any uncertainty in the system is nominal, due to the intrinsic, small level of error in the radiation readings and the inability to guarantee the absence of shielding material in the lorry. The corresponding output is shown in Figure 5.

Amber / alerts potentially caused by benign materials - Truck B Truck B sets off the radiation alarms. The manifest contains the phrases "Bananas for Tesco's from Sri Lanka, Terracotta tiles for Topps tiles from Italy, medical equipment for NHS Greater London, from Siemens in Germany". The corresponding evidence tree is shown in Figure 6.

The amber alert is triggered because the overall uncertainty computed at the top node is 0.236 (greater than the threshold of 0.1 for a "green" light). The decision tree breaks down the decision so as to show where the concern lies. The manifest indicates that the source itself is safe, whereas the scanners have detected radiation.

Figure 6: Sample decision tree computed for Truck B.

Red - suspicious Truck C Truck C does not set off any alarms. However, the manifest merely reads "grouped delivery from Ukraine". The operator is unable to figure out what is in the cargo. The recent occurrence of radioactive smuggling instances in the Ukraine, and the lack of any specific source within the Ukraine also raises suspicions. A brief visual inspection indicates that there is a lot of scope for shielding materials.

The corresponding evidence tree is shown in Figure 7. The top node shows an aggregated uncertainty of 0.445, This is especially due to uncertainty about the content. The balance of probability mass is weighted to the worst end of the scale, indicating that there are plenty of grounds for suspicion. Although the alarms were not triggered, there were lots of potential shielding materials, which add to uncertainty in the reading, but also lead to very poor reliability of the readings.

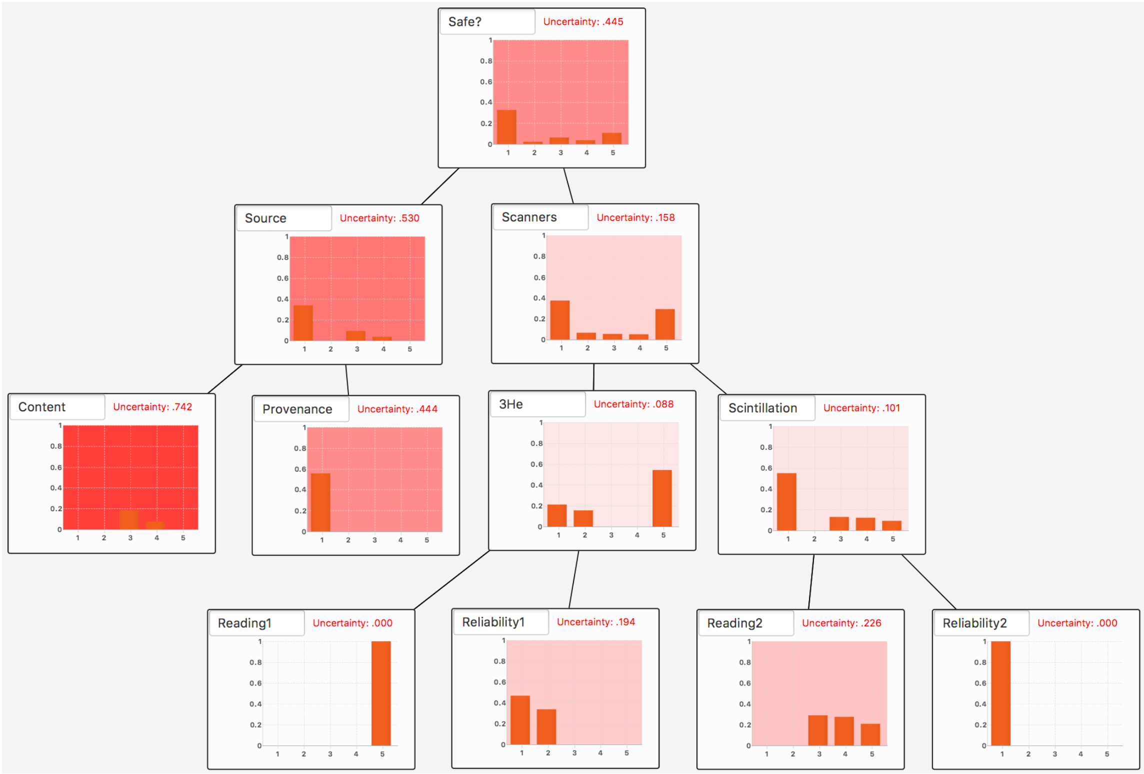


Figure 7: Sample decision tree computed for Truck C.

4 Conclusions

The project has shown that it is feasible to apply ER for radiation detection. The key benefits include:

- The ability to *explicitly* represent doubt or uncertainty, arising from a lack of information.
- The ability to 'operationalise' decision making – to remove variation that arises due to subjective opinions of operators, to set out the precise decision processes that are to be adopted.
- Built-in explanations of *why* a particular detection decision has yielded the result that it has.

5 Future Work

We have provided an idea of how data can be entered for a relatively specific scenario. However, this has primarily been for the purpose of illustration. Ultimately, the technique needs to be straightforward to use, and there are strong drivers to remove as much effort from the operator as possible.

To this end, one fruitful area of future work would be to develop techniques by which to pre-process manifest data. If, for example, the contents of a cargo can be automatically extracted from a manifest, this removes the burden from the operator. Accordingly, one area of future work will be to apply text-processing techniques, such as Topic Mining, to manifest data, with the aim of automatically identifying contents and origins.

Another important direction is the broad decision structure. This would of course need to be refined with input from operators themselves.

Importantly, one aspect that has not featured in this project is the question of *weighting* different nodes differently. This feature can, for example, be used to "switch off" particular sub-trees of a decision tree, or to at least add levels of precedence. These weights could either be 'hard-coded' in, or manipulated by the operator during operation.

References

[NWK14] Sunil Nair, Neil Walkinshaw, and Tim Kelly. Quantifying uncertainty in safety cases using evidential reasoning. In *International Conference on Computer Safety, Reliability, and Security*, pages 413–418. Springer, 2014.

[NWKdlV15] Sunil Nair, Neil Walkinshaw, Tim Kelly, and Jose Luis de la Vara. An evidential reasoning approach for assessing confidence in safety evidence. In *Software Reliability Engineering (ISSRE), 2015 IEEE 26th International Symposium on*, pages 541–552. IEEE, 2015.

[TL15] Nicholas Tsoulfanidis and Sheldon Landsberger. *Measurement and Detection of Radiation*. CRC Press, 2015.

[YX02] Jian-Bo Yang and Dong-Ling Xu. On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 32(3), 2002.