2014 Twelfth Annual Conference on Privacy, Security and Trust (PST)

Communicating and Visualising Multicriterial Trustworthiness under Uncertainty

Florian Volk*, Sascha Hauke*, Daniel Dieth, and Max Mühlhäuser*
*Technische Universität Darmstadt/CASED
Telecooperation Lab
Hochschulstraße 10, 64289 Darmstadt, Germany
florian.volk@cased.de, sascha.hauke@cased.de, daniel.dieth@gmail.com, max@informatik.tu-darmstadt.de

***Abstract*—Visualisations are often used to communicate trustworthiness to end users. Showing a number of stars, for example, is a well-known practice in e-commerce applications to communicate the quality of a product or service. Many products or services also have their quality – in terms of trustworthiness – described along more than one dimension, so that not only an overall trust score has to be communicated, but multiple scores, one for each dimension. Current visualisations of such a multicriterial trustworthiness are often based on the display of multiple individual star-like interfaces – a practice that offers room for improvement with regard to intuitive understanding of the displayed trust information. In this paper, we present *T-Viz*, a trust visualisation based on radar plots and pie charts. *T-Viz* concurrently shows multiple trust scores, one for each dimension, along with an aggregated trust score. Moreover, *T-Viz* also shows a reliability measure for every trust score graphically, in the form of a certainty score. The evaluation results from a pilot study with eleven participants indicate that *T-Viz* is an intuitive, comprehensible and clear interface. It succeeds at visualising and communicating multicriterial trust scores under uncertainty in one, easy to understand, graphical representation.**

I. Introduction

Trust and reputation systems (TRS) have become a widely used tool for guiding and securing user decisions in a multitude of application scenarios. In order to properly function as decision support mechanisms for human users, for instance in e-commerce scenarios, the trust or reputation values computed by a TRS have to be conveyed to the user in a manner which allows for both intuitive and quick comprehension, as well as analytical clarity. This need is exacerbated by the fact that in many scenarios not just one trust or reputation value has to be conveyed, but rather a combination of such values.

Consider the example of comparing two hotels on a hotel booking site, as depicted symbolically in Fig. 1. Here, the user is typically presented with an aggregate overall score for each hotel, as well as component scores for a number of different sub-categories, such as *cleanliness*, *noise*, and *service*. Often, the number of sub-categories in hotel booking sites is even larger than the three sub-categories shown symbolically in Fig. 1: *agoda.com* gives four, *hotels.com* gives five, *expedia.com*, *hrs.com* and *tripadvisor.com* distinguish six, and *hotel.de* provides seven sub-categories. Aggregating these different sub-categories into a weighted or unweighted cumulative value carries the risk of information loss. For instance, in Fig. 1 it is obvious that the two hotels differ in sub-category quality despite having the same overall trust score. Therefore, a visualisation method for trust scores with multiple sub-categories (to which we will refer as *multicriterial trust scores*) should be able to transport the complete information of the sub-categories with as little loss as possible.

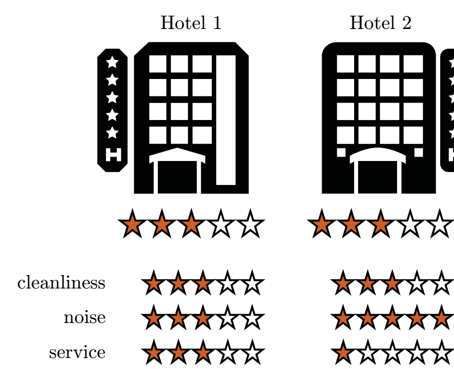

Fig. 1. Two hotels with an average rating of 3/5 stars each. However, only the ratings for the individual criteria show the differences between the hotels.

However, this in itself is still not sufficient. TRS compute trust values that are typically statistical point estimates of future behaviour. These estimates may be interpreted as probability estimates of the outcome of a future interaction. As such, trust scores or, in general, statements on trust, carry with them an inherent uncertainty (compare [1], [2]). On statistical grounds, it is therefore necessary to supply a measure of the assumed accuracy of the trust estimate. State-of-the-art trust models [3] account for this by supplying an (un-)certainty score along with a trust score, while commercial reputation systems typically only display the number of ratings that lead to a particular score – if they do so at all.

In a comprehensive visualisation for multicriterial trust scores, all of these aspects have to be combined: an ability to clearly, intuitively and in an analytically sound way capture the trust scores of multiple categories, as well as depicting the concordant certainty of each of the trust scores. In this paper, we present and evaluate a novel trust visualisation, *T-Viz*. *T-Viz* is designed to display trust scores for all the criteria of a multicriterial trust score along with their associated uncertainty values as well as an aggregated overall trust score.

T-Viz combines radar charts, which are well-known two-dimensional visualisations for multivariate data, with pie charts, resulting in a propeller-style visualisation of multicriterial trust scores as shown in Fig. 2. Each slice represents the expected quality of one criterion or sub-category. The height of a slice denotes the trust score, its width the level of certainty. The slice's color is an aggregation of trust and certainty, calculated as per Ries' *CertainTrust* [4] expectation value computation, and maps this value linearly to an RGB colour gradient from red to green. Thus, the color already is a per-criterion aggregation of a trust score into an intuitively graspable colour scheme. Furthermore, an overall trust score is displayed as percentage value in the center of the *T-Viz* chart. Within the scope of this paper, this overall, central trust score is the weighted average of the trust scores of all criteria.

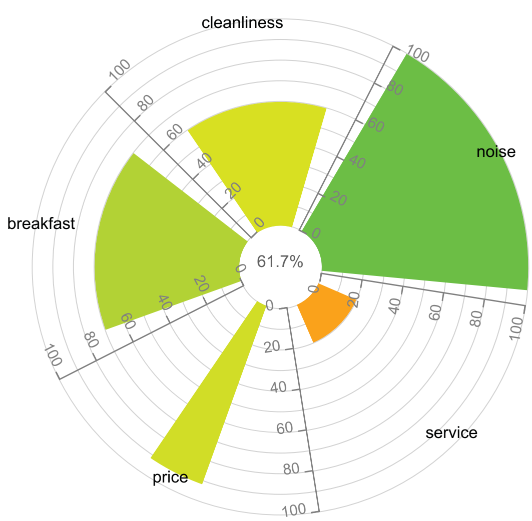

Fig. 2. A T-Viz chart displaying exemplary data for a hotel.

In order to show its efficacy, the *T-Viz* chart was evaluated in a pilot study with eleven users. Results from this first study (see Sect. IV) are promising and indicate that *T-Viz* provides improved performance with regard to intuitive reception and speed of decision-making. We are currently conducting a larger study with an online questionnaire to verify the results of the pilot study.

The remainder of this paper is structured as follows: Sect. II reviews relevant related work that influenced the development of *T-Viz*. A detailed description of the *T-Viz* chart and a documentation of the design process can be found in Sect. III. Sect. IV presents the evaluation of *T-Viz* and discusses results from the pilot study with eleven participants.

II. Related Work

The most widely applied visualisations of trustworthiness use variations of the well-known stars interface to indicate a trustworthiness value. For example, Amazon uses up to five stars to display the customer rating of a seller in its marketplace[1]. After a transaction, the customer can rate the seller in various criteria with one to five stars each. However, potential customers are presented with a seller's overall score instead of scores for each category. As an overall confidence measure in the stars estimate, the total number of ratings that went into the trust score is also presented. Various hotel booking and rating sites, to give another example, also provide similar visualisations, displaying a star rating for the overall score and one for each sub-category (exemplified in Fig. 1).

The *T-Viz* visualisation of multicriterial trust scores is fundamentally rooted in multicriterial data representation based on radar plots. The visualisation of multicriterial aspects of trust, presented by Nurse et al. in [5] has the same foundation as their trustworthiness visualisation is also based on radar graphs. They define five *trust factors*—namely competence, proximity, popularity, recency, and corroboration—and plot their individual values on the axes of a pentagon-shaped radar graph, as shown in Fig. 3. Moreover, they highlight the area formed by connecting the points where the values are drawn on the axes. Their work was supported by a study with 40 participants. This study showed that the radar graph-style visualisation's size as well as the order and distribution of the five axes may influences the user's perception of the plotted overall trust. In comparison with the propeller-style visualisation presented in the paper at hand, the visualisation from [5] displays exactly five fixed criteria and does not visualise uncertainty assigned to the values of these criteria.

The opinion triangle by Jøsang et al. displays a single binomial opinion in a triangular barycentric coordinate system with the dimensions belief, disbelief, uncertainty, and base rate [6] (cf. Fig. 4). The opinion triangle takes uncertainty into account. While the opinions on several criteria can be aggregated into one opinion using Subjective Logic, only one opinion can be displayed with the opinion triangle at a time. It is of course possible to display each sub-category in an individual opinion triangle. The opinion triangle is also very well suited to the analytically exact display of trust scores; it is however, not readily accessible on an intuitive level.

Similar in capabilities to the opinion triangle is Ries' Human Trust Interface [7]. It can display one opinion (which might be an aggregation of several opinions) and shows the corresponding uncertainty. Moreover, the Human Trust Interface uses a red-to-green colour gradient to indicate the trustworthiness of the displayed opinion (cf. Fig. 5). Still, the Human Trust Interface is limited to showing one opinion at a time. The use of colours of the visualisation in the paper at hand is equivalent to the one from the Human Trust Interface. By representing the expectation value as colour value, the Human Trust Interface also trades off quantitative-analytical capabilities for qualitative-intuitive comprehension.

Kelly et al. conducted a study on the effectiveness of food labelling with 790 participants [8]. Among others, they compared a multicriterial visualisation based on the colouring of traffic lights with one that additionally shows an aggregated score alongside the scores for the individual criteria (cf. Fig. 6). While both of these visualisations significantly outperformed the other tested visualisations, there was no significant difference between these two in terms of reading speed or understandability. Nevertheless, we decided to include an aggregated score into the *T-Viz* visualisation for those users that only need a fast overview instead of detailed information on the underlying criteria.

Based on the results of Kelly et al., Idris et al. developed the Colour Coded Traffic Light Labelling as an indicator of trustworthiness for user-generated content in digital maps [9].

[1] https://www.amazon.com/feedback

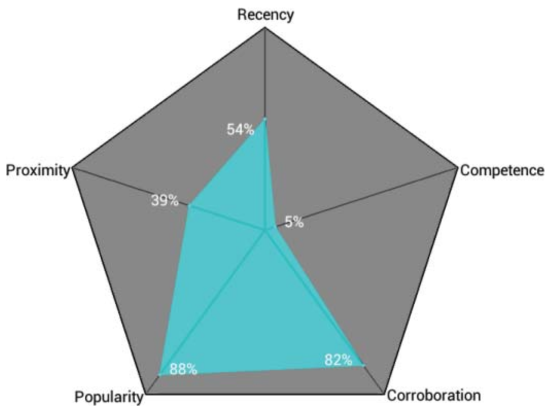

Fig. 3. Multicriterial visualization using a radar chart by Nurse et al. [5]. The black background of the figure has been removed for better display.

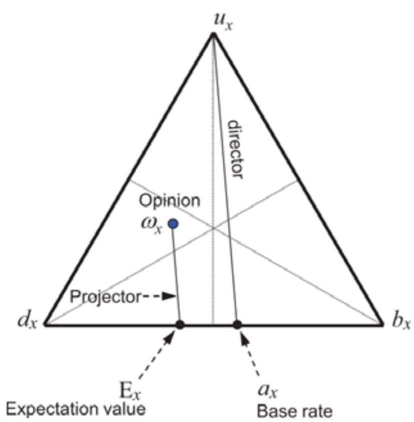

Fig. 4. Opinion Triangle by Jøsang et al. [6].

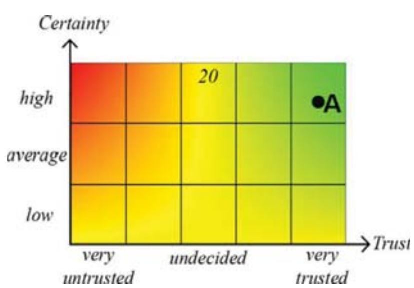

Fig. 5. Human Trust Interface by Ries et al. [7].

Fig. 6. Traffic Light + Overall Rating for front-of-pack food labelling by Kelly et al. [8]. The original figure from the referenced paper was set in levels of grey.

Hence, their visualisation had a fixed set of criteria and only the overall trustworthiness is encoded in a red, yellow or green coloured circle along with a textual description. We decided to display the aggregated score in our visualisation as percentage value to also accommodate achromatopsiac viewers.

III. Visual Presentation and Design Decisions

In this section (III-A) we explain the manner in which trust data is displayed using *T-Viz* charts. Additionally, the three major iterations during the design process of T-Viz are documented (III-B).

A. Visual Presentation

The *T-Viz* chart is intended to show several independent criteria or sub-categories of trust simultaneously. We consider a criterion to be a more abstract dimension in trust formation than a sub-category; in this sense, a criterion is akin to one of the trust factors presented in Nurse et al. [5], while a sub-category is a concrete dimension, such as the rating for the cleanliness of a hotel room. For all intents and purposes, we will use the terms sub-category and criterion interchangeably, as the *T-Viz* chart is applicable to both. In order to display multiple criteria or sub-categories, we partition a circular area into segments. Each segment is dedicated to one criterion and is labeled with the criterion's name. The segments are separated by axes for better readability.

Within a segment, a coloured slice of a circle visualises the trust and certainty scores for a criterion. The slice's height denotes the trust value while at the same time, the slice's width denotes the certainty associated with the trust value. In an analogy to traffic lights, the colour of the slice describes a score derived from both trust and certainty. Possible colours range in a gradient scale from pure green (high expected quality/trust) gradually over yellow to pure red (low expected quality/trust). The RGB-values of the slice's colour are determined the same way as Ries' Human Trust Interface calculates its color gradient [7]. The certainty values presented in the tested implementation were computed according to [7], using the *CertainTrust* SDK[2]. Consequently, certainty is a straightforward function of the number of ratings, n, for each sub-category, so that certainty $c(n) = \frac{n}{n+2}$.

In the center of the *T-Viz* chart, an overall score is shown as percentage value with one decimal place. Here, the overall score is the weighted average of all criteria displayed in the segments. Other ways of computing the overall score are possible, but are relegated to future work, together with other modes of visualising it, such as colouring the center.

In the implementation, *T-Viz* charts offer interactive features, as well. Hovering a mouse pointer over a slice shows both trust and certainty in numerical values. This interactive feature was not used in the study presented in Sect. IV.

B. Design Iterations

The current design of *T-Viz* charts is the result of the three major iterations (cf. Fig. 7 (a)–(c)), which will be

[2]Available at: *http://www.tk.informatik.tu-darmstadt.de/de/research/smart-security-and-trust/*

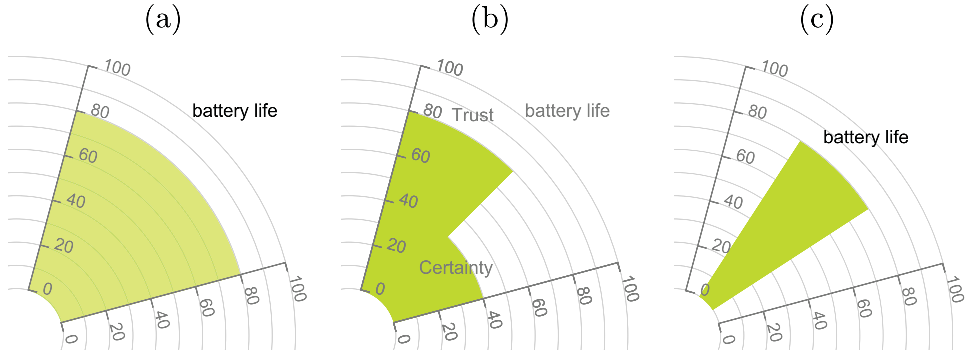

Fig. 7. Evolutionary steps during the design of T-Viz: (a) height denotes trust while opacity denotes certainty, (b) independent slices for trust and certainty, (c) height denotes trust while width denotes certainty.

described in the following. While the basic idea and shape of *T-Viz* charts did not change, the representation of trust and certainty was iteratively changed over the development process to accommodate user comments.

1) First Iteration (a): In the first designs as shown in Fig. 7 (a), a slice's height indicates the trust value in that criterion. In contrast to later design, the certainty is denoted by the opacity of the slice. This approach leads to problems with very low values for certainty – in these cases, the slice itself is barely visible. In these cases, the highly transparent colours of the slice are also hard to differentiate with regard to the displayed trust score, so that the user is unable to see if a trust score is high (green) or low (red).

2) Second Iteration (b): As a consequence from the findings with the first iteration, the second one eliminates variations in the opacity entirely, as shown in Fig. 7 (b). Trust and certainty are displayed as two adjoining slices in the same segment. The slice with the lower angular minute arc denotes trust while the other one denotes certainty. To differentiate between the two values, both slices are labeled.

This design has been found to have two considerable drawbacks. Both slices have the same colour. The closer the trust value and the certainty value, the more difficult it becomes to differentiate both slices. Even though a differentiation is not necessary anymore in such cases, feedback from users indicated that they are distracted by this behaviour. In an intermediary iteration, both slices were coloured independently according to their absolute value. This modification made trust and certainty scores readily distinguishable. However, the labelling for two slices in conjunction with the label for the criterion's name caused users to comment on an "overcrowded look" of this representation. Especially for high values of trust and certainty, the labels overlap and thus, tend to become difficult to read and, hence, confusing.

3) Current Iteration (c): In an attempt to overcome the drawbacks of the first two iterations, the current version of *T-Viz* uses only one unlabelled slice per criterion, the height and width of which denote trust and certainty concurrently, as shown in Fig. 7 (c). The slice's colour represents an aggregated value for the whole segment. This iteration of *T-Viz* was evaluated in a pilot study, the design and results of which are described in the next section.

IV. Evaluation

In order to evaluate how well the *T-Viz* chart performs in conveying multicriterial trust information, we compared it against a multicriterial stars interface in a limited user study with eleven participants. The stars interface was chosen because it is widely encountered in existing e-commerce sites and, as a consequence, has achieved a high degree of familiarity among users.

In order to confirm that the *T-Viz* chart is intuitive and provides quick comprehension, we hypothesise that when tasked with choosing between two alternatives based on (a subset) of multi criterial trust scores

1) users will choose the more highly trusted alternative using *T-Viz* at least as often as when using the stars interface, and
2) users will make their choice as fast as or faster than the stars interface when using *T-Viz*.

A. Pilot Study

For testing the aforementioned hypotheses, we have designed a user study that will be described in the following.

1) Study Setup: In a computer-based user study, *T-Viz* charts were tested against the stars interface with regard to their efficacy in respect to quality and speed of decision. Initially, each participant was given a brief introduction to both the stars interface and the *T-Viz* chart in the form of a written and illustrated explanation of the interface. Next, the participants were given the task of choosing the "better" of two functionally equivalent products or services. For this, they were presented with the task of selecting the alternative that exhibits the better trust scores along a given subset of sub-categories or that exhibits the better overall trust score. For example, participants were given the following task: "*From the following two alternatives, select the* better designed *and* cheaper product*! Make your choice solely based on the supplied trust information!*" For this example, the presented trust information for each product consisted of an overall trust score, as well as a trust score for the following sub-categories: *price*, *design*, *workmanship*, and *ease-of-use*.

Each participant was presented with thirteen distinct pairs of equivalent services or products *twice*, once displaying the trust information in stars interfaces and once displaying it in

T-Viz charts. Thus, each participant had to make 26 individual choices from thirteen pairings. The sequence and order of the products and visualisation types was randomised throughout the user study in order to account for memory effects. For each pair, the time it took the participant to make a decision was recorded, as was the choice of alternative.

In case of displaying *T-Viz* charts, the participants were presented with two charts per pairing, i.e., one per alternative. Each chart was divided into sub-categories in the same manner, thus displaying the same sub-categories for each alternative, in the same order. The trust and certainty values represented in the two charts were, of course, dependent on the sub-category of the individual alternative, and could, hence, be different between the two alternatives. When displaying the information in a stars interface representation, the participants were shown a number of 5-star interfaces (similar to the prototypical representation in Fig. 1): one 5-star interface for the overall score, and one per sub-category. The stars were filled in half-star steps. Two of such groupings of five-star interfaces were presented per pairing, one per alternative.

Additionally, for four selected services/products, the participants were asked whether they preferred the *T-Viz* or stars interface visualisations, and to map the presented trust scores to a numerical rating given in percentages. Finally, qualitative feedback on general preferences regarding the two visualisations and free form comments were elicited from the participants.

Each participant was given a brief introduction to the computer-based survey tool used to conduct the study. After this orientation, no further instructions were given to the participant. Each participant completed the study by him- or herself without interaction with the supervising personnel or other participants.

2) Participants: For the pilot study, we recruited eleven participants, two female and nine male. The participants were between 23 and 50 years old (see, Fig. 8). All participant reported to be frequent users of e-commerce, ten reporting to use an e-commerce site 3 to 4 times a month on average, one participant used e-commerce sites once a month on average. Consequently, all participants also reported a high to very high familiarity with the stars interface,

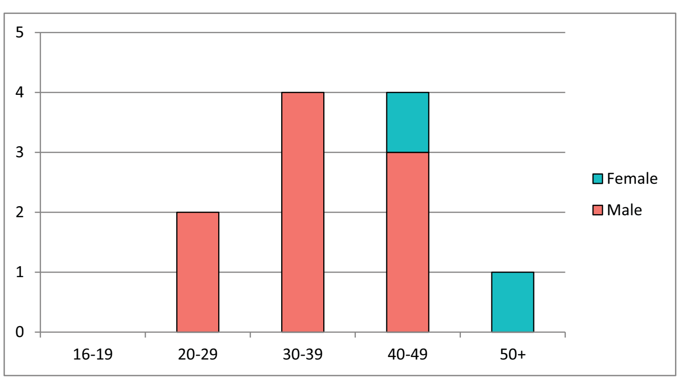

Fig. 8. Age and gender distribution of pilot study participants.

B. Results and Discussion

Quantitative analysis of the results of the pilot study was centred around accepting or rejecting the two hypotheses introduced in Sect. IV. For testing hypothesis H_1: *"users will choose the more highly trusted alternative using* T-Viz *at least as often as when using the stars interface"*, we compared the chosen alternative in each of the 13 pairings of equivalent products/services to the correct choice that *should* have been made by the participants. For each pairing such a correct choice existed and was defined as a higher cumulative number of stars or a higher cumulative trust score over the relevant sub-categories.

As a result of the pilot study, hypothesis H_1 could not be rejected. There was a non-significant performance advantage of the *T-Viz* chart that became slightly more pronounced as the number of displayed sub-categories increased (from four to seven). Overall, the participants selected the correct alternative – as an average over all 13 pairings – 92.31% of the time (standard deviation 12.22%) using the stars interface. Using *T-Viz* charts, the average percentage of correct selection was 95.8% (standard deviation 4.72%).

In terms of correctness of selection, the *T-Viz* chart is thus not inferior to the stars interface. In fact, it performs slightly better, particularly when the number of sub-categories increases. This indicates that the *T-Viz* chart provides more clarity than the stars interface.

In order to test hypothesis H_2: *"users will make their choice as fast as or faster than the stars interface when using* T-Viz*"*, we measured the time it took each participant to choose one of the two alternatives per pairing. As shown in Fig. 9, decision times were considerably lower when using *T-Viz* instead of the stars interface.

Because a *Shapiro-Wilk* test [10] showed that the time differences between the decision times using the stars interface and *T-Viz* could not be assumed to follow a Gaussian distribution, a *Wilcoxon-Mann-Whitney* test [11] was used to test hypothesis H_2. The one-sided test for significance confirmed H_2: when using the *T-Viz* chart, the participants were significantly faster (at a p-value of $p < 0.001$) than when using the stars interface. An *ANOVA* and *t-test* were also run on the data and confirmed H_2 at high significance levels but were not relied upon because, technically, they require Gaussianity of the data.

We take the results of the significance tests to be indicative of the qualities we desired of the novel *T-Viz* charts. In spite of the participants high familiarity with the stars interface, the *T-Viz* chart performed consistently better in sub-category driven selection tasks. Although the study size of only eleven participants does not permit us to pass *final* judgement, the results are highly promising that *T-Viz* achieves the goal of providing intuitive and quick comprehension, as well as analytical clarity.

In addition to the quantitative results, the pilot study also yielded qualitative feedback. The participants stated almost unequivocally that they would prefer the *T-Viz* interface for selections that require more than 4 sub-categories to be displayed. For a smaller number of sub-categories, the stars interface was either preferred by the majority of participants

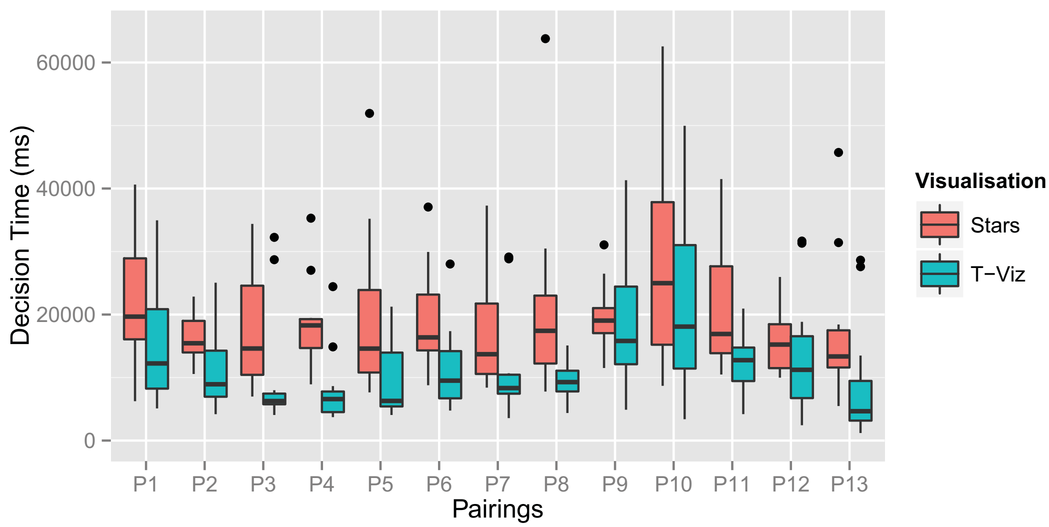

Fig. 9. Time taken by participants to choose between equivalent alternatives (in milliseconds): Stars interface vs. *T-Viz* chart.

($\leq$ 3 sub-categories, 36% in favour of *T-Viz*, 64% in favour of the stars interface) or there was a nearly even split (4 sub-categories, 55% in favour of *T-Viz*, 45% in favour of the stars interface).

When asked their opinions on what they would consider strengths and weaknesses of *T-Viz*, participants pointed out that one *T-Viz* chart is very easy to compare to another, in terms of visual clarity, and facilitates the identification of the respective advantages and shortcomings of the services or products that are rated in the chart. Particularly for personally subjective sub-categories, *T-Viz* was considered to have a '*more intuitive feel*', as one of the participants stated. There was also a noticeable trend among participants to remark upon the unexpected ease-of-use of the *T-Viz* chart. Several participants stated that the explanation of the *T-Viz* chart before the study gave them the impression that *T-Viz* was complex and difficult to grasp and that they were surprised when this turned out not to be the case.

Among the weaknesses that should be improved upon, participants suggested a larger, possibly colour-coded area for the overall score.

V. Conclusions and Future Work

The results obtained from our pilot study are promising indicators that improvements in the visualisation of multi-criterial trust scores, under assumptions of uncertainty, are possible without sacrificing ease-of-use and user acceptance. The presented interface, the *T-Viz* chart, is also an – in our opinion important – evolutionary development of the multi-criterial display of trust scores in radarplots, as proposed by Nurse et al. [5]. For instance, *T-Viz* charts are largely invariant to the ordering of the sub-categories within the chart, due to its propeller-shaped design, which is slice-centric instead of axis-centric, as it is in radarplots. Thereby, *T-Viz* charts avoid a change in the covered area when reordering the sub-categories – a possible point of confusion for users.

The pilot study will be extended in the future into an Internet-based full study. It has prompted several conceptual improvements that will be integrated into the online study, for instance a modified and more highly visible overall trust score. We are confident that the first results from the pilot study can be confirmed and made statistically more relevant by recruiting more study participants in the future.

As a practical application, the integration of *T-Viz* charts with databases requiring the display of multicriterial trust information is already ongoing work. Specifically, the Cloud Security Alliance's *Security, Trust and Assurance Registry* (STAR)[3] is currently being used as a data source for a *T-Viz* based demonstrator, extending work begun in [12].

In conclusion, the results in the paper at hand present a very strong indication that intuitive, comprehensible and clear interfaces for visualising and communicating multicriterial trust scores under uncertainty are possible – in the form of the *T-Viz* chart. In the pilot study, the *T-Viz* chart has generally outperformed the widely used stars interface, in spite of a very high familiarity of the participants with the latter interface. While the sample of participants will need to be considerably expanded in a future study, that also encompasses further interface types and has more statistical power, the work presented in this paper provides a sound foundation for further strides in the chosen direction. We are also hopeful that this work will cross-fertilise other, related areas within the field of computational trust, such as the visualisation of *multinomial* trust scores, a highly related, but slightly different topic.

Acknowledgment

The work presented in this paper was performed in the context of the Software-Cluster project *InDiNet* (www.software-cluster.org) and the Software Campus project *MoVe4Dynamic*; it was partially funded by the German Federal Ministry of Education and Research (BMBF) under grant no. "01IC10S04" and grant no. "01IS12054". The authors assume responsibility for the content.

[3] *https://cloudsecurityalliance.org/star/*

References

[1] A. Jøsang, "A Logic for Uncertain Probabilities," *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, vol. 9, no. 3, pp. 279–311, 2001.

[2] D. Gambetta, "Can We Trust Trust?" *Trust: Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford*, pp. 213–237, 2000.

[3] A. Jøsang, R. Ismail, and C. Boyd, "A Survey of Trust and Reputation Systems for Online Service Provision," *Decision Support Systems*, vol. 43 (2), pp. 618–644, 2007.

[4] S. Ries, "Extending Bayesian Trust Models Regarding Context-dependence and User Friendly Representation," in *Proceedings of the 2009 ACM Symposium on Applied Computing*, ser. SAC '09. ACM, 2009, pp. 1294–1301.

[5] J. R. Nurse, I. Agrafiotis, S. Creese, M. Goldsmith, and K. Lamberts, "Communicating Trustworthiness using Radar Graphs: A Detailed Look," in *The 11th International Conference on Privacy, Security and Trust (PST)*. IEEE, 2013.

[6] A. Jøsang, T. Azderska, and S. Marsh, "Trust Transitivity and Conditional Belief Reasoning," in *IFIPTM*, ser. IFIP Advances in Information and Communication Technology, vol. 374. Springer, 2012, pp. 68–83.

[7] S. Ries and D. Schreiber, "Evaluating User Representations for the Trustworthiness of Interaction Partners," in *International Workshop on Recommendation and Collaboration (ReColl'08) in conjunction with the International Conference on Intelligent User Interfaces (IUI'08)*. ACM Press, 2008.

[8] B. Kelly, C. Hughes, K. Chapman, J. C.-Y. Louie, H. Dixon, J. Crawford, L. King, M. Daube, and T. Slevin, "Consumer testing of the acceptability and effectiveness of front-of-pack food labelling systems for the Australian grocery market," *Health Promot Int*, vol. 24, no. 2, pp. 120–9, 2009.

[9] N. H. Idris, M. J. Jackson, and R. J. Abrahart, "Colour Coded Traffic Light Labeling: A Visual Quality Indicator to Communicate Credibility in Map Mash-up Applications," in *International Conference on Humanitites, Social Sciences, Science & Technology (ICHSST 2011)*, 2011.

[10] S. S. Shapiro and M. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52(3-4), pp. 591–611, 1965.

[11] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.

[12] S. M. Habib, S. Ries, M. Mühlhäuser, and P. Varikkattu, "Towards a Trust Management System for Cloud Computing Marketplaces: using CAIQ as a trust information source," *Security and Communication Networks Journal*, 2013.