Supporting Expert Assessment of Argument Structures in Trust Cases

Łukasz Cyra

Janusz Górski

Information Assurance Group
Department of Software Engineering
Gdańsk University of Technology, Poland
Contents

• What is ‘trust case’?
• The Trust-IT framework
• Example argument
• The appraisal scale
• Appraisal example
• The aggregation mechanism
• Conclusions
Trust vs Trustworthiness

• Trust
  - trust is the notion referring to a belief in some postulated property of a trusted object considered in a specific context

• Trustworthiness
  - Trustworthiness is the notion referring to the justification explaining why we should trust that the object exhibits the posulated property in this context

• Trustworthiness can imply Trust
Trust Case

Trust Case is an argument that provides a satisfactory (from a selected viewpoint) justification for a specified set of properties to make a judgement about the trustworthiness of the chosen object.

Trust Case integrates argumentation with the evidence that supports this argumentation.

The notion of Trust Case is a generalization of the common notion of Safety Case.
Language for representing trust cases

Claim

Counter-Argument Strategy

Argument Strategy

Warrant

Assumption

Claim

Fact

Premises

Reference

Claim

Fact

Reference

Counter-Argument Strategy

Argument Strategy

Warrant
Trust Case example

Safety of ANGEL user
- Argument by considering protection against safety hazards
  - WARRANT: All unacceptable identified safety hazards are dealt with
- Correctness of alarms
- Exercise status information and exercise improving comments
- Correctness of user’s data
  - Argument by describing system features
    - WARRANT: Features are located in two layers - platform and application
  - Support from ANGEL platform to correctness of data
    - Argument by describing supporting functionalities and properties
      - WARRANT: Platform maintains reliability of data
      - Reliability of ANGEL data
    - Reliability of data storage and processing of ANGEL application
      - Correct exercise support from ANGEL application
- ANGEL system installation, configuration and maintenance
- Accidental damage of sensors e.g. due to Patient’s physical activity
  - Argument by reference to platform component property
    - WARRANT: Sensor design protects them sufficiently from accidental damage
    - Sensor node resilience
Trust case for ANGEL final demonstrator

- ANGEL system trustworthiness
  - Argument by analysing trustworthiness aspects
    - W WARRANT: Argument from the constitution of essential aspects
  - Safety of ANGEL user
    - Argument by considering protection against safety hazards
      - W WARRANT: All unacceptable identified safety hazards are dealt with
        - Argument by referring to completeness of system hazard analysis
          - W WARRANT: Safety hazard analysis process identified major hazards within the demonstrator scope
          - F Hazard analysis employed a well-defined process derived from standards
          - F Analysis of safety hazards for Angel application scope identified major accidents
            - ANGEL Platform Definition - description of application scope

- Scenario 1
  - New Node
  - View Reference
  - Haz
  - Correct

- REFERENCE
  - Hazard identification for Scenario 1
    - Label:
      - Include in report
    - State: Initial
Language for representing trust cases

Scenarios for using trust cases

Trust case processes

Templates Patterns

Trust case system

Trust-IT framework
Problem

How to assess the ‘strength’ of the argument in a trust case and how to communicate it to the relevant stakeholders

Solution

Provide an argumentation appraisal mechanism which starts from assessments of the facts and inferences in the argument and aggregates them to an assessment of the topmost claim
Trust Case example

- Safety of ANGEL user
  - Argument by considering protection against safety hazards
    - WARRANT: All unacceptable identified safety hazards are dealt with
  - Correctness of alarms
  - Exercise status information and exercise improving comments
  - Correctness of user's data
    - Argument by describing system features
      - WARRANT: Features are located in two layers - platform and application
    - Support from ANGEL platform to correctness of data
      - Argument by describing supporting functionalities and properties
        - WARRANT: Platform maintains reliability of data
    - Reliability of ANGEL data
  - Reliability of data storage and processing of ANGEL application
    - Correct exercise support from ANGEL application
  - ANGEL system installation, configuration and maintainance
  - Accidental damage of sensors e.g. due to Patient's physical activity
    - Argument by reference to platform component property
      - WARRANT: Sensor design protects them sufficiently from accidental damage
  - Sensor node resilience
The appraisal scale

CONFIDENCE SCALE

lack of confidence
with very low confidence
with low confidence
with high confidence
with very high confidence
for sure

DECISION SCALE

rejectable
opposable
tolerable
acceptable
disbelief
belief

uncertainty
Appraisal example
Trust case for ANGEL final demonstrator

- ANGEL system trustworthiness
  - Argument by analysing trustworthiness aspects
    - WARRANT: Argument from the constitution of essential aspects
  - Safety of ANGEL user
    - Argument by considering protection against safety hazards
      - WARRANT: All unacceptable identified safety hazards are dealt with
      - Argument by referring to completeness of system hazard analysis
        - WARRANT: Safety hazard analysis process identified major hazards within the demonstrator scope
    - Hazard analysis employed a well-defined process derived from standards
    - Analysis of safety hazards for Angel application scope identified major accidents
  - Hazard analysis covers the scope of Final Demonstrator

- ANGEL Platform Definition - description of application scope

- Correctness of alarms

**Confidence level:**
- for sure

**Decision:**
- acceptable

Delete assessment
Argument by analysing trustworthiness aspects

WARRANT: Argument from the constitution of essential aspects

Safety of ANGEL user

Argument by considering protection against safety hazards

WARRANT: All unacceptable identified safety hazards are dealt with

 Arguments by referring to completeness of system hazard analysis

WARRANT: Safety hazard analysis process identified major hazards within the demonstrator scope

Hazard analysis employed a well-defined process derived from standards

Analysis of safety hazards for ANGEL application scope identified major accidents

ANGEL Platform Definition - description of application scope

Hazard identification for Scenario 1

Hazard identification for Scenario 2 & 3

System safety hazards assessment report

Hazard analysis covers the scope of Final Demonstrator

Confidence level: with very high confidence

Decision: tolerable
Argument by analysing trustworthiness aspects

WARRANT: Argument from the constitution of essential aspects

Safety of ANGEL user

Argument by considering protection against safety hazards

WARRANT: All unacceptable identified safety hazards are dealt with

Argument by referring to completeness of system hazard analysis

WARRANT: Safety hazard analysis process identified major hazards within the demonstrator scope

Hazard analysis employed a well-defined process derived from standards

Safety and privacy risk assessment process description

Analysis of safety hazards for Angel application scope identified major accidents

Hazard analysis covers the scope of Final Demonstrator

CL Correctness of alarms

CL Exercise status information and exercise improving comments

CL Correctness of user's data

Description Notes Change history Links to node Trust evaluation

Assessor mode Viewer mode

Confidence level:

- with very high confidence

Decision:

acceptable

Apply Cancel
Argument by analysing trustworthiness aspects

WARRANT: Argument from the constitution of essential aspects

Safety of ANGEL user

Argument by considering protection against safety hazards

WARRANT: All unacceptable identified safety hazards are dealt with

Argument by referring to completeness of system hazard analysis

WARRANT: Well-defined safety hazard analysis process identifies all major hazards

F Hazard analysis employed a well-defined process derived from standards

F Analysis of safety hazards for Angel application scope identified major accidents

F Hazard analysis covers the scope of Final Demonstrator

CL Correctness of alarms

CL Exercise status information and exercise improving comments

CL Correctness of user's data

CL ANGEL system installation, configuration and maintenance

Description Notes Change history Links to node Trust evaluation

Assessor mode Viewer mode

Confidence level:
with very high confidence

Decision:
acceptable
Argument by analyzing trustworthiness aspects

**WARRANT:** Argument from the constitution of essential aspects

**Safety of ANGEL user**

**F** Argument by considering protection against safety hazards

**W** Warrant: All unacceptable identified safety hazards are dealt with

**F** Argument by referring to completeness of system hazard analysis

**W** Warrant: Well-defined safety hazard analysis process identifies all major hazards

**F** Hazard analysis employed a well-defined process derived from standards

**F** Analysis of safety hazards for Angel application scope identified major accidents

**F** Hazard analysis covers the scope of Final Demonstrator

**CL** Correctness of alarms

**CL** Exercise status information and exercise improving comments

**CL** Correctness of user’s data

**CL** ANGEL system installation, configuration and maintainance
Argument by analyzing trustworthiness aspects

WARRANT: Argument from the constitution of essential aspects

Safety of ANGEL user

Argument by considering protection against safety hazards

WARRANT: All unacceptable identified safety hazards are dealt with

Argument by referring to completeness of system hazard analysis

WARRANT: Well-defined safety hazard analysis process identifies all major hazards

Hazard analysis employed a well-defined process derived from standards

Analysis of safety hazards for Angel application scope identified major accidents

Hazard analysis covers the scope of Final Demonstrator

ANGEL Platform Definition - description of application scope

Correctness of alarms

Exercise status information and exercise improving comments

Correctness of user's data

Confidence level:

with low confidence

Decision:

tolerable

User name: TCT administrator
Confidence level: with low confidence
Decision: tolerable

Apply  Cancel
Argument by analysing trustworthiness aspects

WARRANT: Argument from the constitution of essential aspects

Safety of ANGEL user

Argument by considering protection against safety hazards

WARRANT: All unacceptable identified safety hazards are dealt with

Argument by referring to completeness of system hazard analysis

WARRANT: Well-defined safety hazard analysis process identifies all major hazards

Hazard analysis employed a well-defined process derived from standards

Analysis of safety hazards for Angel application scope identified major accidents

Hazard analysis covers the scope of Final Demonstrator

ANGEL Platform Definition - description of application scope

Correctness of alarms

Exercise status information and exercise improving comments

Correctness of user's data

Confidence level:

with low confidence

Decision:

tolerable
The aggregation mechanism
Different argument types depending on how the premises contribute to the conclusion

Different aggregation rule for each argument type

Mapping of the linguistic values on Dempster-Shaffer belief and plausibility functions

A-argument rule
Yager’s modification of Dempster’s rule of combination

\[
Bel(c) = Bel(a_1) \cdot Bel(a_2) + Bel(a_1) \cdot (Pl(a_2) - Bel(a_2)) + Bel(a_2) \cdot (Pl(a_1) - Bel(a_1))
\]

\[
Pl(c) = 1 - (1 - Pl(a_1))(1 - Pl(a_2)) + (1 - Pl(a_1))(Pl(a_2) - Bel(a_2)) + (1 - Pl(a_2))(Pl(a_1) - Bel(a_1))
\]
**Validation experiment**

- **30 students involved**
- **Repeated assessment**

<table>
<thead>
<tr>
<th>Aggregation rule</th>
<th>Consistency of students’ assessments</th>
<th>A</th>
<th>a₁</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Confidence scale</strong></td>
<td><strong>Decision scale</strong></td>
<td><strong>Confidence scale</strong></td>
</tr>
<tr>
<td><strong>A-rule</strong></td>
<td>1,03</td>
<td>0,64</td>
<td>1,04</td>
</tr>
<tr>
<td><strong>NSC-rule</strong></td>
<td>0,94</td>
<td>0,53</td>
<td>1,06</td>
</tr>
<tr>
<td><strong>C-rule</strong></td>
<td>0,84</td>
<td>0,88</td>
<td>0,91</td>
</tr>
</tbody>
</table>
Conclusion

• Trust-IT provides for development, maintenance and sharing of trust cases for real life objects
  - A Personalized Information Platform for health and life Services
    - (6th EU FR Integrated Project PIPS)
  - A platform supporting WSN based health related applications
    - (6th EU FR STREP Project ANGEL)
  - TTA based dependable embedded systems
    - (6th EU FR Integrated Project DECOS)
  - Support for standards conformance (e.g. ISO 27001, ISO 14971:2000)
  - Trustworthiness of HON (Helth On the Ne) criteria

• Argument appraisal mechanism provides for third party assessment of trust cases

• Linguistic scales support communication of trust case contents between stakeholders

• More experiments are needed to calibrate and validate the appraisal mechanism