Graphical vs. Tabular Notations for Risk Models: On the Role of Textual Labels and Complexity

Katsiaryna (Kate) Labunets
TU Delft, The Netherlands
E: k.labunets@tudelft.nl

Joint work with
Fabio Massacci, University of Trento, Italy
Alessandra Tedeschi, DeepBlue srl, Rome, Italy
Rationale

• Risk recommendations should be “consumed” mostly by not-experts in security

• What if the security representation is not easy to understand?
  – Stakeholder does not understand you
  – The security recommendations are not implemented

• “Understand” !≠ “Believe to have understood”
Example Risk Models

A simple example of one attack path represented in graphical and tabular notation.

**CORAS diagram**
- Customer shares credentials with next-of-kin
- Customer
  - Threat: Lack of compliance with terms of use
  - Vulnerability: Customer shares credentials with next-of-kin
- Customer
  - Treatment: Regularly inform customers of terms of use
- Unauthorized account login [unlikely]
  - Likelihood: Unlikely
  - Consequence: Unauthorized account login
  - Asset: Integrity of account data

**UML-style diagram**
- Customer shares credentials with next-of-kin
- Customer
  - Threat: Lack of compliance with terms of use
  - Vulnerability: Customer shares credentials with next-of-kin
- Customer
  - Treatment: Regularly inform customers of security best practices
  - Consequence: Unauthorized account login [Likelihood: unlikely]
  - Asset: Integrity of account data

**NIST table row entry**

<table>
<thead>
<tr>
<th>Threat event</th>
<th>Threat source</th>
<th>Vulnerability</th>
<th>Impact</th>
<th>Overall likelihood</th>
<th>Level of impact</th>
<th>Asset</th>
<th>Security control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer shares credentials</td>
<td>Customer</td>
<td>Lack of compliance with terms of use</td>
<td>Unauthorized account</td>
<td>Unlikely</td>
<td>Severe</td>
<td>Integrity of account data</td>
<td>Regularly inform customers of terms of use</td>
</tr>
<tr>
<td>with next-of-kin</td>
<td></td>
<td></td>
<td>login</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Previous work

• Published in EMSE journal:

• Included two studies:
  – [2014] 69 MSc and BSc students from Italy and Brazil
  – [2015] 83 professional post-master course and MSc students from Italy

• Treatment
  – graphical and tabular risk modeling notations

• Findings
  – Tabular is more effective that the graphical representation for simple comprehension tasks
  – Less difference for complex tasks, but still tabular is better
Research questions

We address the following questions for participants with a significant work experience:

**RQ1**: Does the task complexity have an impact on the comprehensibility of the models?

**RQ2**: Does the availability of textual labels improve the participants effectiveness in extracting correct information about security risks?
Experiment Description [1/2]

• Goal:
  – Compare tabular vs. graphical risk models w.r.t. comprehensibility

• Treatments
  – Notations: NIST 800-30 (tabular); CORAS (graphical); UML-style (graphical)
  – Task: open questions with different level of complexity about information presented in the model
    • 7 questions (originally 12 but 5 questions were discarded due to an implementation error)

• Between-subject design
  – one treatment per participant
Experiment Description [2/2]

- **Application scenario:**
  - Online Banking scenario by Poste Italiane

- **Recruitment process:**
  - Email invitation distributed via mailing lists by UNITN and DeepBlue
  - Offered a reward of 20 euro (via PayPal)
  - 572 attempts to start the experiment

- **Participants:**
  - 61 professional (avg. 9 years of working experience)

The number of participants reached each experimental phase

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Consented</th>
<th>Provided demographics</th>
<th>Finished task</th>
<th>Total valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tabular</td>
<td>39</td>
<td>30</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>UML</td>
<td>30</td>
<td>23</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>CORAS</td>
<td>40</td>
<td>29</td>
<td>21</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>82</td>
<td>63</td>
<td>61</td>
</tr>
</tbody>
</table>
Demographics

Age

- 24–30 yrs old: 36%
- 31–40 yrs old: 41%
- 41–62 yrs old: 23%

Education degree

- BSc: 45%
- MSc: 36%
- MBA: 8%
- PhD: 11%

Working experience

- Did not report: 3%
- Had 1-3 yrs: 18%
- Had 4-7 yrs: 43%
- Had >7 yrs: 36%

Experience in graph. modeling languages

- Novices: 16%
- Beginners: 2%
- Competent users: 23%
- Proficient users: 31%
- Experts: 28%
### Used Risk Models: NIST

<table>
<thead>
<tr>
<th>Threat Event</th>
<th>Threat Source</th>
<th>Vulnerabilities</th>
<th>Impact</th>
<th>Asset</th>
<th>Overall Likelihood</th>
<th>Level of Impact</th>
<th>Security Controls</th>
</tr>
</thead>
</table>
| Customer’s browser infected by Trojan and this leads to alteration of transaction data | Hacker           | 1. Poor security awareness 2. Weak malware protection | Unauthorized transaction via web application | Integrity of account data | Likely             | Severe          | 1. Regularly inform customers about security best practices.
2. Strengthen authentication of transaction in web application. |
| Keylogger installed on computer and this leads to sniffing customer credentials. Which leads to unauthorized access to customer account via web application. | Cyber criminal   | Insufficient detection of spyware              | Unauthorized transaction via web application | Integrity of account data | Likely             | Severe          | Strengthen authentication of transaction in web application. |
| Spear-phishing attack on customers leads to sniffing customer credentials. Which leads to unauthorized access to customer account via web application. | Cyber criminal   | Poor security awareness                        | Unauthorized transaction via web application | Integrity of account data | Likely             | Severe          | 1. Regularly inform customers about security best practices.
2. Strengthen authentication of transaction in web application. |
| Keylogger installed on customer’s computer leads to sniffing customer credentials | Cyber criminal   | Insufficient detection of spyware              | Unauthorized access to customer account via web application | User authenticity       | Certain             | Severe          | Regularly inform customers about security best practices. |
| Spear-phishing attack on customers leads to sniffing customer credentials    | Cyber criminal   | Poor security awareness                        | Unauthorized access to customer account via web application | User authenticity       | Certain             | Severe          | Regularly inform customers about security best practices. |
| Keylogger installed on customer’s computer leads to sniffing customer credentials | Cyber criminal   | Insufficient detection of spyware              | Unauthorized access to customer account via web application | Confidentiality of customer data | Certain             | Severe          | Regularly inform customers about security best practices. |
| Spear-phishing attack on customers leads to sniffing customer credentials    | Cyber criminal   | Poor security awareness                        | Unauthorized access to customer account via web application | Confidentiality of customer data | Certain             | Severe          | Regularly inform customers about security best practices. |
| Fake banking app offered on application store and this leads to sniffing customer credentials | Cyber criminal   | Lack of mechanisms for authentication of app | Unauthorized access to customer account via fake app | User authenticity       | Likely             | Critical         | Conduct regular searches for fake apps.               |
| Fake banking app offered on application store and this leads to sniffing customer credentials | Cyber criminal   | Lack of mechanisms for authentication of app | Unauthorized access to customer account via fake app | Confidentiality of customer data | Likely             | Severe          | Conduct regular searches for fake apps.               |
| Fake banking app offered on application store leads to sniffing customer credentials. Which leads to unauthorized access to customer account via fake app. | Cyber criminal   | Lack of mechanisms for authentication of app | Unauthorized access to customer account via fake app | Confidentiality of customer data | Unlikely            | Minor           | Conduct regular searches for fake apps.               |
| Fake banking app offered on application store leads to alteration of transaction data | Cyber criminal   | Lack of mechanisms for authentication of app | Unauthorized transaction via Poste App         | Integrity of account data | Unlikely            | Minor           | Conduct regular searches for fake apps.               |
| Smartphone infected by malware and this leads to alteration of transaction data | Hacker           | Weak malware protection                        | Unauthorized transaction via Poste App         | Integrity of account data | Unlikely            | Minor           | Regularly inform customers about security best practices. |
| Denial-of-service attack                                                      | Hacker           | 1. Use of web application 2. Insufficient resilience | Online banking service goes down              | Availability of service  | Certain             | Minor           | 1. Monitor network traffic.
2. Increase bandwidth. |
| Web-application goes down                                                     | System failure   | Immature technology                            | Online banking service goes down              | Availability of service  | Certain             | Minor           | Strengthen verification and validation procedures.    |
Used Risk Models: CORAS
Comprehension Questions

We ask to identify a risk element of a specific type that is related to another element of a different type.

“Which threats can exploit the vulnerability ‘Poor security awareness’? Please specify all threats:”

At least one question per element type:

Graphical element types:
1. Threat
2. Vulnerability
3. Threat scenario
4. Unwanted incident
5. Likelihood
6. Consequence
7. Asset
8. Treatment

Tabular element types:
1. Threat source
2. Vulnerability
3. Threat event
4. Impact
5. Overall likelihood
6. Level of impact
7. Asset
8. Security control
Comprehension Questions: Complexity

• Task complexity is based on [Wood, 1986].
• Our mapping of Wood’s components to the elements of modeling notations:
  – **Information cues (IC)** describe some characteristics that help to identify the desired element of the model.
    • Which are the assets that can be harmed by the unwanted incident “Unauthorized access to HCN”?
  – **Required acts (A)** are judgment acts that require to select a subset of elements meeting some explicit or implicit criteria.
    • What is the highest consequence?
  – **Relationships (R)** between a desired element and other elements of the model that must be identified.
    • … the assets that can be harmed by Cyber criminal…

\[
\text{Complexity (Q}_i\text{)} = |\text{IC}_i| + |\text{R}_i| + |\text{A}_i|
\]
Comprehension Questions: Complexity

• Simple question:
  – “What are the assets that can be harmed by Cyber criminal?”

• Complex question:
  – “Which unwanted incidents may be exploited by Hacker to harm the asset “Integrity of account data” with the highest likelihood?”

Legend:
- Information cue
- Relationship
- Required act
Measurements

- **Precision** of the response to a question:
  - \( \frac{\text{# of identified correct elements}}{\text{# of all elements in a response}} \)

- **Recall** of the response to a question:
  - \( \frac{\text{# of identified correct elements}}{\text{# of all expected correct elements}} \)

- **F-measure** is a weighted harmonic mean of precision and recall
  \[
  F\text{-measure} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}
  \]

- **Subject’s Comprehension**
  - Aggregated F-measure of all questions about assigned risk model
Results

All questions

<table>
<thead>
<tr>
<th>METHOD</th>
<th>T</th>
<th>UML</th>
<th>CORAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tabular</td>
<td>4</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>UML</td>
<td>1</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>CORAS</td>
<td>1</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Median all = 0.91

Aggregated Recall vs Aggregated Precision

Median all = 0.83
Results: RQ1 – Complexity effect

- F-measure of simple (compl. lvl =2) vs. complex (compl. lvl >2) questions:
  - UML & CORAS: Small difference, but not statistically significant
  - Tabular: Statistical equivalence using an equivalence test (TOST)

Table: F-measure by task complexity

<table>
<thead>
<tr>
<th></th>
<th>Simple</th>
<th>Complex</th>
<th>W</th>
<th>TOST_W</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>med</td>
<td>sd</td>
<td>mean</td>
</tr>
<tr>
<td>Tabular</td>
<td>0.94</td>
<td>1.00</td>
<td>0.15</td>
<td>0.91</td>
</tr>
<tr>
<td>UML</td>
<td>0.84</td>
<td>1.00</td>
<td>0.29</td>
<td>0.78</td>
</tr>
<tr>
<td>CORAS</td>
<td>0.66</td>
<td>0.67</td>
<td>0.31</td>
<td>0.63</td>
</tr>
</tbody>
</table>

No effect has been observed in our experiment
Results: RQ2 – Textual labels

Table: Precision and recall by modeling notation

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>med</td>
</tr>
<tr>
<td>Tabular</td>
<td>0.92</td>
<td>1.00</td>
</tr>
<tr>
<td>UML</td>
<td>0.83</td>
<td>0.91</td>
</tr>
<tr>
<td>CORAS</td>
<td>0.68</td>
<td>0.75</td>
</tr>
</tbody>
</table>

- **Precision:**
  - Tabular and UML are equivalent
    - $p_{\text{TOST-MW}} = 0.04$
  - Tabular > CORAS
    - $p_{\text{MW}} = 0.0009$, $p_{\text{KW}} = 0.002$

- **Recall:**
  - Tabular > UML
    - $p_{\text{MW}} = 0.004$, $p_{\text{KW}} = 0.008$
  - Tabular > CORAS
    - $p_{\text{MW}} = 1.4 \cdot 10^{-5}$, $p_{\text{KW}} = 6.6 \cdot 10^{-5}$
  - UML > CORAS
    - $p_{\text{MW}} = 0.04$

Availability of textual labels helps to elicit better responses
RQ2: Discussion

- Tables are good, but textual labels also help
  - Tables provide possibility to use computer-aided search and copy&paste information
  - CORAS group used search feature more often than UML group
- CORAS does not have textual labels to identify elements of necessary type
Conclusions

- Electronic tables may be your best choice to communicate security risk assessment results with stakeholders.
- Pure graphical models are prone to comprehension errors (comparing to tables).

Mitigation options:
- Textual labels [cheap]
- Invest more in training stakeholders on notation [expensive]
Open questions

• How well models support memorization of information about security risk
  – What information you can recall from your memory?

• Task complexity
  – Can we measure it better?
  – Which questions are complex enough to trigger declared benefits of graphical models?