PRELIMINARY EXPERIMENTS ON RELATIVE COMPREHENSIBILITY OF TABULAR & GRAPHICAL RISK MODELS

Katsiaryna Labunets
University of Trento, Italy (katsiaryna.labunets@unitn.it)

Joint work with
Yan Li¹, Fabio Massacci², Federica Paci³, Martina Ragosta⁴, Bjørnar Solhaug¹, Ketil Stølen¹, Alessandra Tedeschi²

¹SINTEF, ²University of Trento, ³University of Southampton, ⁴DeepBlue
Motivation - 1

- Risk recommendations should be “consumed” mostly by not-experts in security
- Security Risk Assessment in ATM
  - SESAR SecRAM method
    - Tabular-based
    - **Non-experts** in security can apply it
  - Future methods
    - new graphical models to support risk assessment
Motivation - 2

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

• Goal
  • Tabular vs. graphical risk models: which is easier to understand?

• Treatments
  • Graphical risk model (CORAS)
  • Tabular risk model (NIST)

• Context
  Security risk assessment for the Online Banking scenario
  • Participants
    • 35 MSc students – University of Trento, Italy
    • 11 MSc students – University of Oslo, Norway
  • 8 comprehensibility question
Risk Modeling: Tables vs. Diagrams

CORAS diagram

<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 with next-of-kin</td>
<td>Customer</td>
<td>Lack of compliance with terms of use</td>
<td>Unauthorized account login</td>
<td>Unlikely</td>
<td>Severe</td>
<td>Integrity of account data</td>
<td>Regularly inform customers of terms of use</td>
</tr>
</tbody>
</table>
Used Risk Models: CORAS
## 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>
<tbody>
<tr>
<td>Customers' browser infected by Trojan and this leads to Hackers altering transaction data.</td>
<td>Hacker</td>
<td>1. Poor security awareness 2. Weak malware protection</td>
<td>Unauthorized transaction via web application.</td>
<td>Integrity of account data</td>
<td>Likely</td>
<td>Severe</td>
<td>1. Regularly inform customers about security best practices to increase their security awareness and improve malware protection. 2. Strengthen authentication of transaction in web application to prevent unauthorized transaction via web application.</td>
</tr>
<tr>
<td>Keylogger installed on computer and this leads to sniffing customer credentials. Which leads to unauthorized access to customer account via web application.</td>
<td>Cyber criminal</td>
<td>Insufficient detection of spyware</td>
<td>Unauthorized transaction via web application.</td>
<td>Integrity of account data</td>
<td>Likely</td>
<td>Severe</td>
<td>1. Regularly inform customers about security best practices to increase their security awareness. 2. Strengthen authentication of transaction in web application to prevent unauthorized transaction via web application.</td>
</tr>
<tr>
<td>Spear-phishing attack on customers leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Poor security awareness</td>
<td>Unauthorized transaction via web application.</td>
<td>Integrity of account data</td>
<td>Likely</td>
<td>Severe</td>
<td>1. Regularly inform customers about security best practices to increase their security awareness. 2. Strengthen authentication of transaction in web application to prevent unauthorized transaction via web application.</td>
</tr>
<tr>
<td>Keylogger installed on customer's computer and this leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Insufficient detection of spyware</td>
<td>Unauthorized access to customer account via web application.</td>
<td>User authenticity</td>
<td>Certain</td>
<td>Severe</td>
<td>Regularly inform customers about security best practices to increase their security awareness.</td>
</tr>
<tr>
<td>Spear-phishing attack on customers leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Poor security awareness</td>
<td>Unauthorized access to customer account via web application.</td>
<td>User authenticity</td>
<td>Certain</td>
<td>Severe</td>
<td>Conduct regular searches for fake apps.</td>
</tr>
<tr>
<td>Keylogger installed on customer's computer leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Insufficient detection of spyware</td>
<td>Unauthorized access to customer account via web application.</td>
<td>Confidentiality of customer data</td>
<td>Certain</td>
<td>Severe</td>
<td>Conduct regular searches for fake apps.</td>
</tr>
<tr>
<td>Spear-phishing attack on customers leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Poor security awareness</td>
<td>Unauthorized access to customer account via web application.</td>
<td>Confidentiality of customer data</td>
<td>Certain</td>
<td>Severe</td>
<td>Conduct regular searches for fake apps.</td>
</tr>
<tr>
<td>Fake banking app offered on application store and this leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Lack of mechanisms for authentication of app</td>
<td>Unauthorized access to customer account via fake app.</td>
<td>User authenticity</td>
<td>Likely</td>
<td>Minor</td>
<td>Conduct regular searches for fake apps.</td>
</tr>
<tr>
<td>Fake banking app offered on application store and this leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Lack of mechanisms for authentication of app</td>
<td>Unauthorized access to customer account via fake app.</td>
<td>Confidentiality of customer data</td>
<td>Likely</td>
<td>Minor</td>
<td>Conduct regular searches for fake apps.</td>
</tr>
<tr>
<td>Fake banking app offered on application store and this leads to sniffing customer credentials.</td>
<td>Cyber criminal</td>
<td>Lack of mechanisms for authentication of app</td>
<td>Unauthorized transaction via Fake App.</td>
<td>Integrity of account data</td>
<td>Unlikely</td>
<td>Minor</td>
<td>Regularly inform customers about security best practices to improve malware protection of their smartphones.</td>
</tr>
<tr>
<td>Smartphone infected by malware and this leads to alteration of transaction data by Cyber criminal.</td>
<td>Cyber criminal</td>
<td>Weak malware protection</td>
<td>Unauthorized transaction via Fake App.</td>
<td>Integrity of account data</td>
<td>Unlikely</td>
<td>Minor</td>
<td>1. Monitor network traffic to prevent use of web application by hackers. 2. Increase bandwidth to improve resilience.</td>
</tr>
<tr>
<td>Denial-of-service attack.</td>
<td>Hacker</td>
<td>1. Use of web application 2. Insufficient resilience</td>
<td>Online banking service goes down.</td>
<td>Availability of service</td>
<td>Certain</td>
<td>Minor</td>
<td>Strength verification and validation procedures to prevent use of immature technology.</td>
</tr>
<tr>
<td>Web-application goes down</td>
<td>System failure</td>
<td>Immature technology</td>
<td>Online banking service goes down.</td>
<td>Availability of service</td>
<td>Certain</td>
<td>Minor</td>
<td></td>
</tr>
</tbody>
</table>
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:”

One question per element type:

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

- **Precision** of the response to a question:
  - # of identified correct elements / # of all listed elements

- **Recall** of the response to a question:
  - # of identified correct elements / # 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**
  - Average F-measure of all questions about assigned risk model
Experimental Protocol

• Training
  • Training on both risk modeling notations [8 min]
  • General introduction to the application scenario [2 min]
  • Demographics & Background questionnaire [5 min]

• Application
  • Comprehension questionnaire [20 min]
    • 8 questions
  • Post-task questionnaire [2 min]
    • To control possible effect of the experimental settings on the results

• Evaluation
  • 2 researchers independently checked the subjects’ responses against the predefined set of correct answers
Data Collection

- Between subject design
  - One subject received only one of two risk models
- 24 subjects were discarded
  - Due to incorrect time limit in SurveyGizmo
- In total we got data from 22 subjects
  - Tabular: 13 subjects
  - Graphical: 9 subjects
Preliminary Results

All questions (Q1–Q8)

Distribution of mean precision and recall per subject by risk model type

- **Tabular**
  - T: N = 1
  - G: N = 0
  - **Median all = 0.91**

- **Graphical**
  - T: N = 5
  - G: N = 3
  - **Median all = 0.83**
  - T: N = 6
  - G: N = 4

**Average Precision vs. Average Recall**
**Preliminary Results**

- [Overall] Tabular = Graphical
  - 10% better mean recall using tabular risk model
    - => more complete responses

<table>
<thead>
<tr>
<th>Mean</th>
<th>Tabular</th>
<th>Graphical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.9</td>
<td>0.88</td>
</tr>
<tr>
<td>Recall</td>
<td>0.87</td>
<td>0.79</td>
</tr>
<tr>
<td>F-measure</td>
<td>0.89</td>
<td>0.83</td>
</tr>
</tbody>
</table>

- Need replications
  - At least 116 subjects in total for F-measure
Threats to validity

• **Internal validity**
  • Search in the risk model
    • Tabular: 62% of subjects used search (only 1 subject in Oslo)
    • Graphical: 22% of subjects used search

• **External validity**
  • Participants are students
    • We will replicate study with professionals
  • Only CORAS and NIST
    • Need to study other representations

• **Conclusion validity**
  • Statistical power
    • We plan to replicate the study
Summary

• Conclusions
  • Which representation is better?
    • Participants’ level of comprehension is the same
  • Tables showed 10% better recall
    • More complete response → less chance to overlook things

• Future work
  • Replication with more subjects (professionals and students)
  • Different risk modeling notations
  • Task complexity factor

• Ads
  • Want to join the effort? → we are looking for replications
  • More Info? → http://securitylab.disi.unitn.it