Security Literature Review Paper Analysis Rubric
Technical Report Number SERG-2015-01

Amiangshu Bosu, Chris Corley, Jeffrey Carver*, Matthias Gander,
Jason King, Sedef Kocak, and Jouni Markkula

*Department of Computer Science
University of Alabama
carver@cs.ua.edu

August 4, 2015

Abstract

This document contains the definition of a rubric used to classify security research papers. First we define three dimensions used to classify each paper: a) what is being analyzed in the paper - the evaluation subject, b) whether the evaluation subject is new, meaning whether it was first proposed in this paper or not, and c) how the authors evaluated the properties of the evaluation subject - the evaluation approach. Second, we provide a methodology for categorizing each paper by mapping the three dimensions into one string. Finally, for each evaluation approach, we then define a series of questions a reviewer can answer to help determine the completeness of the report, from a Science of Security perspective.

1 Dimensions

This section defines the three dimensions used to characterize each paper. Note that there is a many-to-many relationship between the *evaluation subject* and the *evaluation approach*. That is, there could be multiple *evaluation subjects* in each paper and each *evaluation subject* could have multiple *evaluation approaches*.

1.1 Evaluation Subject

The item being evaluated in the paper (a paper could have more than one of these).

M - *Model* - graphical or mathematical description of a system and its properties.

L - *Language* - a programming language.

PL - *Protocol* - A written procedural method that specifies the behavior for data exchange amongst multiple parties.

PR - *Process* - computational steps to transform one thing into something else.

T - *Tool* - an implementation of a process.

AL - *Algorithm/Theory*- Proposes a new algorithm or update to an existing algorithm or a new theory.

1.2 Evaluation Approach

The approach used to evaluate the *evaluation subject*. Each *evaluation subject* will likely have one or more of these.

EX - *Experiment* - An orderly process that seeks to test a hypothesis, generally has 2 or more treatments (e.g. experimental, control, baseline, or some other means to countermeasure confounding variables) and seeks to evaluate the effect of the treatment on the subjects of the study. Experiments are differentiated from Case Studies in that the subject pool is typically large enough to test the hypotheses statistically.

CS - *Case Study* - Tested on one or more real or example systems. A case study is differentiated from an EX in that a case study is a detailed, in-depth examination of one (or a small number) of cases relative to the *evaluation subject* [1].

Q - *Survey* - A set of questions (questionnaire/interview/focus group/ opinion poll) aimed at gathering data from people regarding the *evaluation subject*.

P - *Proof* - A formal or mathematical process to show that the evaluation subject is true or correct.

D - *Discussion/Argumentation* - Discussion, opinions, or argumentation regarding the evaluation subject without providing a proof or empirical data (note, this category does not refer to a discussion of the results obtained by some other method of evaluation. It only includes papers in which the only evaluation is Discussion/Argumentation).

1.3 IsNew

The *evaluation subject* (i.e., model/language/protocol/algorithm/process/tool) may be new (i.e., the authors are proposing) or existing. This dimension defines, for each *evaluation subject*, whether it was proposed in this paper or proposed elsewhere, as follows.

N - The *evaluation subject* is first proposed (described) in this paper.

E - The *evaluation subject* was proposed (described) in another paper.

2 Mapping

When using this rubric for a group of researchers to analyze a large set of papers, it is helpful to have a shorthand method for recording the dimensions described in Section 1. This section describes an approach for efficiently recording this information to ease the process of comparing results.

- Map the *evaluation subjects* directly to the *evaluation approaches*, separating multiple *evaluation subjects* with an ampersand (&). Also, for each *evaluation subject* put the value for **IsNew** in square brackets - [ ]
  - M[E]; P & PR[N]; CS
    - Here, we are explicitly stating that P relates only to M which was an Existing Model, and CS relates only to a New PR
    - Each evaluation subject[IsNew] should have it's own set of evaluation approach(es).

- If the *evaluation approach* should not be split, use a slash (/):
  - M[E]; CS/P & PR[N]; CS
    - Here, P relates only to M, but CS was used for both M and PR. "killing 2 birds with 1 stone"
- If it is unclear that the *evaluation approach* should be split, please denote that specific *evaluation approach* with an asterisk:
  - M[E]; CS*/P & PR[N]; CS
    - Here, it is unclear if the CS is related to the both PR and M, or just PR. It may be, but the paper does not describe the difference well enough to say.
    - It would be best to try to resolve these during discussion.

If you're into BNFs, heres a grammar:
multimap = map | map '&' multimap
map = evaluationSubject[IsNew] ';' evaluationApproaches
evaluationApproaches = maybe_evaluationApproach | maybe_evaluationApproach '/' evaluationApproach
maybe_EvaluationApproach = evaluationApproach — evaluationApproach '*'
evaluationSubject[IsNew] = 'M[E]' | 'PR[N]' | . . .
evaluationApproach = 'CS' | 'EX' | 'D' | . . .

3 Rubric Questions

For each *evaluation approach* defined in Section 1 this section provides a number of rubric questions that can be answered to help evaluate the completeness of the report. Each rubric questions can be answered as *Yes*, *No*, or *Partial* (as defined in the rubrics that follow).

- Yes means "the information is present in the paper and easy to find (i.e. well-formatted)"
- Partially means "the information is present in the paper but may not be easy to find"
- No means "the information is omitted from the paper"

In most cases, we drew on published guidelines in building these rubrics. The citation next to each *evaluation approach* indicates the source from which we drew information in building that particular rubric.

3.1 Experiment [2]

EX1 : Are the research objectives described? (e.g., goals, questions, hypotheses)?

Yes - Clearly defined a labeled (e.g. Research Question, RQ, Objective,)

Partial - Included in the text but not clearly labeled

No - Not present

EX2 : Are the methods for subject sampling described? (e.g., recruitment/selection process)?

Yes - Explicitly defined in the text

No - Not defined in the text

EX3 : Are the data collection procedures described (e.g. definition of the metrics/variables, operational constructs, measurement levels)?

Yes - Explicitly described in the text

No - Not described in the text

EX4 : Are the analysis procedures described? (e.g., hypothesis checks, statistical tests, p-values, performance metrics, precision, recall, accuracy, False positive, False negative etc.)?

Yes - Paper includes all of the following: statistical tests (by name) or other analysis method, results of statistical test (including p-value)

Partial - Paper includes some but not all of the above

No - Paper includes none of the above

EX5 : Are the characteristics of the sample/ systems described? (e.g., demographics, specification)?

Yes - Paper explicitly describes the characteristics of the sample

No - Paper does not explicitly describe characteristics of the sample

EX6 : Does the data presented have descriptive stats? (e.g., mean, std dev, charts or tables to describe data, etc)

Yes - Paper contains a description of the data: e.g., mean/median, standard deviation, frequency, etc...

No - Paper does not describe the data

EX7 : Do they discuss results in relation to the research objectives? (e.g., hypotheses evaluated, questions answered, or "big picture")

Yes - There is a separate discussion section

Partial - The results are discussed, but not in a separate section

No - The results are not discussed

EX8 : Is there a dedicated discussion of the threats to validity (i.e., limitations or mitigations)?

Yes - There is a separate Threats to Validity Section

Partial - Threats to validity are discussed, but not in a separate section

No - Threats to validity are not discussed

3.2 Case Study [4]

CS1 : Are the research objectives described? (e.g., goals, questions, hypotheses)?

Yes - Clearly defined early in the paper (i.e. not in the results or discussion) and labeled (e.g. in bold, italics, underlined or set apart from the text with labels like Research Question, RQ, Objective,)

Partial - Included in the text but either in the wrong location or not clearly labeled (see Yes above)

No - Not present

CS2 : Are the case and its units of analysis described? (i.e., what is the context of the study, what is being tried)?

Yes - The paper explicitly defines the context of the study (i.e. the problem background or why it is important to study these particular research questions or problems) and what is being tried

Partial - The defines some, but not all, of the above

No - The paper defines none of the above

CS3 : Are the methods for subject selection described? (e.g., inclusion/exclusion criteria)?

Yes - The paper explicitly describes how the cases were selected including the rationale for selecting the particular case(s)

No - The paper does not explicitly describe how the cases were selected

CS4 : Are the data collection procedures (i.e., how was this completed) and research instruments (i.e. questionnaire, mining tools, performance computation) described?

Yes - Described in the text

No - Not described in the text

CS5 : Are the analysis procedures described? (e.g., hypothesis checks, statistical tests, p-values, performance metrics, precision, recall, accuracy, False positive, False negative)?

Yes - Paper includes both the statistical tests (by name) or other analysis method (e.g. performance measures) and the results of statistical test (including p-value) or other analysis method

Partial - Paper includes one of the above

No - Paper includes none of the above

CS6 : Is data presented with appropriate descriptive statistics to provide an understanding of the analysis? (e.g., mean, std dev, charts or tables to describe data, etc)?

Yes - Paper contains a description of the data, e.g. mean, median, standard deviation, frequency, charts, tables to describe the data etc

No - Paper does not contain a description of the data

CS7 : Do they discuss results in relation to the research objectives? (e.g., hypotheses evaluated, questions answered, or ”big picture”)?

Yes - There is a separate discussion section

Partial - The results are discussed, but not in a separate section

No - The results are not discussed

CS8 : Is there a dedicated discussion of the threats to validity (i.e., limitations or mitigations)?

Yes - There is a separate Threats to Validity Section or Limitations Section

Partial - Threats to validity are discussed, but not in a separate section

No - Threats to validity are not discussed

3.3 Qualitative Data (Questionnaires, Surveys, Focus Groups, etc...) [5]

Q1 : Are the research objectives described? (e.g., goals, questions, hypotheses)?

Yes - Clearly defined a labeled (e.g. Research Question, RQ, Objective,)

Partial - Included in the text but not clearly labeled

No - Not present

Q2 : Is a rationale behind the questions given? (i.e., why these questions and not others)?

Yes - Each question has a rationale provided

Partial - Some questions have a rationale

No - No questions have a rationale

Q3 : Are the evaluation procedures described?

Yes - The paper explicitly links the research objectives to the survey questions.

No - The paper does not explicitly link the research questions/hypotheses to the survey questions.

Q4 : Are the respondents described? (e.g., demographics)?

Yes - The text explicitly describes the characteristics relevant to the research objectives

No - The text does not describe the relevant characteristics

Q5 : Are the sampling methods described? (i.e., why these respondents? e.g., mailing list, advertised, etc)?

Yes - The paper explicitly describes: the method for recruiting participants, how the questionnaire was advertised, why the participants are the correct ones, inclusion/exclusion criteria

Partial - The paper explicitly describes some, but not all of the above

No - The paper describes none of the above

Q6 : Is how the responses were processed described? (e.g., cleaning data, answer coding)?

Yes - The paper explicitly describes how the data was cleaned, how the answers were coded, triangulation of data and inter-rater reliability (only for qualitative analysis)

Partial - The paper describes some, but not all of the above

No - The paper describes none of the above

Q7 : Is there a dedicated discussion of the threats to validity (i.e., limitations or mitigations)?

Yes - There is a separate Threats to Validity Section

Partial - Threats to validity are discussed, but not in a separate section

No - Threats to validity are not discussed

3.4 Proof [3]

P1 : Is the theorem being proved stated? (i.e., goal)?

Yes - Theorem is explicitly stated

No - Theorem is not explicitly stated

P2 : Are any assumptions used described?

Yes - Assumptions are described

No - Assumptions are not described

P3 : Is informal material given to provide intuition on how the proof works?

Yes - There is informal material, such as a proof sketch or an explanation of the proof in context.

No - There was no sketch or context

P4 : Is where the proof ends marked? (e.g., is there a clear ending of the proof before other, possibly unrelated, text begins)?

Yes - There is a clear end to the proof

No - There is no clear end to the proof

3.5 Discussion

D1 : Is the goal of the argument described?

Yes - The goal of the argument is explicitly described

No - The goal of the argument is not explicitly described

D2 : Are two or more premises and a conclusion given? (Aristotle's rule)?

Yes - Two or more premises and a conclusion are given

Partial - Some, but not all of the above are given

No - None of the above are given

D3 : Is the related knowledge described?

Yes - Related knowledge is explicitly described

No - Related knowledge is not explicitly described

D4 : Is the supporting evidence described or cited?

Yes - Supporting evidence is described or cited

No - Supporting evidence is not described or cited

Acknowledgments

We would like to thank the following people for their reviews of the rubric and their feedback: Ayse Bener (Ryerson University), Michael Felderer (University of Innsbruck), Markku Oivo (University of Oulu), Clemens Sauerwein (University of Innsbruck), and Laurie Williams (North Carolina State University).

References

[1] A. L. George and A. Bennett. *Case studies and theory development in the social sciences.* Mit Press, 2005.

[2] A. Jedlitschka and D. Pfahl. Reporting guidelines for controlled experiments in software engineering. In *International Symposium on Empirical Software Engineering*, page 10, Nov 2005.

[3] L. Lamport. How to write a proof. *American Mathematical Monthly*, 1995.

[4] P. Runeson and M. Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131–164, 2009.

[5] F. J. Shull, J. Singer, and D. I. K. Sjoberg, editors. *Guide to Advanced Empirical Software Engineering.* Springer, 2008.