Security Literature Review Paper Review Rubric
Version 2.0

Morgan Burcham and Jeffrey Carver*

*Department of Computer Science
University of Alabama
carver@cs.ua.edu
August 30, 2016

Abstract

This document contains the definition of a rubric used to classify security research papers. First we define five dimensions used to classify each paper: a) Evaluation Subject – what is being analyzed in the paper, b) Evaluation Subject Source – whether the Evaluation Subject was first introduced in the current paper or elsewhere and by whom, c) Evaluation Attribute – what aspect of the Evaluation Subject is being studied, and d) Evaluation Approach – how the authors evaluated the properties of the Evaluation Subject. For each Evaluation Approach, we define a Completeness Rubric containing a series of questions a reviewer can answer to help determine the completeness of the report from a Science of Security perspective.

1 Introduction

Our work has focused on the use of this rubric in conjunction with security papers to determine the completeness of the information provided in the literature. In order for the security research community to move forward, such literature should contain enough details to aid in scientific tasks such as replication, meta-analysis, and theory building. In the review process, we used the Nvivo data analysis software. We created an Nvivo template for all reviewers. In this template, we created nodes for every item in the rubric. Reviewers would import the security papers into the Nvivo template file. As the reviewers read the papers, they would answer all rubric items by marking the appropriate text and selecting the correct node corresponding to each rubric item.

2 Paper Characterization

This section defines the five dimensions used to characterize each paper. Note that the relationship between evaluation subject and evaluation approach is many-to-many. That is, there could be multiple evaluation subjects in each paper and each evaluation subject could have multiple evaluation approaches. In the case of multiple evaluation subjects, the reviewer will simply mark all subjects that apply and select the corresponding evaluation approaches for each subject. In the case of multiple instances of the same evaluation subject, the reviewer will make a note to that effect on the paper. The reviewer will classify the evaluation approach for each instance of the subject. In the case of multiple instances of the same evaluation approach, the reviewer should make a note to that effect on the paper. For each rubric answer for the evaluation approach, the reviewer will
select the appropriate answer in each instance. For example, if a paper uses multiple Proofs to evaluate a Protocol, the reviewer will mark the individual P1-P4 responses for every proof present in the paper. The reviewer will make note of this special circumstance.

2.1 Evaluation Subject

The item being evaluated in the paper. Note that a paper could have more than one of these. The values for this characteristic are:

M - Model - graphical or mathematical description/representation of a system and its properties. Provides a simplified understanding of a system.

L - Language - a constructed/formal language developed as a method of communication.

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, model, or protocol. An executable piece of software.

TH - Theory - Proposes a new theory or update to an existing theory.

2.2 Evaluation Subject Source

The evaluation subject may be new (i.e., first introduced in the current paper) or existing (i.e., first introduced elsewhere). The values for this characteristic are:

AH - Authors Here: Authors introduced the subject first time in the paper.

AE - Authors Elsewhere: Authors introduced the subject in previous paper.

OM - Other Modified: Someone else introduced the subject and authors modified it.

ON - Other Not Modified: Someone else introduced the subject and authors used it without modification.

2.3 Evaluation Attribute

This characteristic captures which aspect of the evaluation subject is evaluated in the paper. In example, a paper may be evaluating the usability of the evaluation subject. Similarly, a paper may be evaluating some other aspect of the evaluation subject such as the memory usage of the subject. A paper may have multiple evaluation aspects. List all aspects of the evaluation subject which are being evaluated.

O - The categories for this attribute will be built using a Grounded Theory approach based on the data available in the set of papers.
2.4 Evaluation Approach

The approach used to evaluate the *evaluation subject*. Each *evaluation subject* will have one or more of these approaches associated with it.

**E - Empirical** - A process of collecting and analyzing data from a set of participants (who or what is being observed in the study e.g. people, systems, etc...) to determine the distribution of and/or the correlation between variable(s). If the Evaluation Approach is of this type, then it will also need to be characterized with the following three attributes:

- **Identification of Participants** – Determine the participants (source of the data) used in the study.
  
  **SIM** - A special type of participant is a simulation, being the representation of the behavior or characteristics of an evaluation subject through the use of another system, especially a computer program designed for the purpose. This means that the source of the data is coming from a prototype.
  
  **H** - This type of participant is used when humans are the source of the data (i.e. collecting data from interviews, surveys, etc.).
  
  **S** - A system provides data for the study (i.e. benchmarks of system, etc.).

- **Type of Study** – Determine and classify how the study was performed.
  
  **Observational** - Study is performed in a natural setting in which the researcher collects data via observation without intentionally manipulating the environment or behavior of the participants. In this type of study, the researcher is merely observing the participants in a natural setting without interacting with the participants. This includes surveys, being a set of questions (questionnaire, interview, focus group, opinion poll, etc.) aimed at gathering data from human subjects regarding the evaluation subject.
  
  **Interventional** - Researcher intentionally applies treatment(s) to participants that potentially manipulate the participants’ environment or behavior. When multiple treatments are considered, participants are assigned to treatment groups and the effects of the treatments are compared across the groups. One of these treatments could be a ”control” where essentially no intervention is made.

- **Type of Data Gathered** – Determine what type of data was gathered in the study. Multiple types of data may be used in the study.
  
  **Self-reported** - Data consists of self-reported data such as that from interviews, surveys, etc.
  
  **Observed** - The study makes use of recorded observations as its source of data. A researcher observes and collects the data.
  
  **Automated** - The study makes use of data that has been automated in some way (i.e. by a tool, machine, etc.).

- **Number of Study Conditions or Treatments Observed/Measured** – Count of the number of study conditions or treatments included in the study as well as the number of observations taken for each condition or treatment.

- **Number of Subjects** – Count of the number of subjects included in the study.

- **Comparison** – Whether the results from the current study are compared against one or more historical baselines.
H - Historical comparison against old results in a different study.
G - Comparison against generated new data for the same purpose of the study.
N - No comparison at all.

P - Proof - A formal or mathematical process to show that the properties of the evaluation subject are 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).

3 Rubric Questions

For each evaluation approach defined in Section 2 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)"
- Partial 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 Empirical Studies [?, ?]

EM1 : Are the research objectives of the study described? (e.g., goals, questions, hypotheses)?

Yes - Clearly defined and labeled (e.g. Research Question, RQ, Objective, )
Partial - Included in the text but not clearly labeled
No - Not present

EM2 : Is the context of the study described? Does the paper offer details on what is being tried to solve the research problem?

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 paper defines some, but not all, of the above
No - The paper defines none of the above

EM3 : Are the methods for subject sampling described? (e.g., recruitment/selection process, inclusion/exclusion criteria)?

Yes - Explicitly defined in the text
No - Not defined in the text
EM4: Are the data collection procedures (e.g., how was this completed, definition of the metrics/variables, operational constructs, measurement levels) and research instruments (i.e. questionnaire, mining tools, performance computation) described?

Yes - Explicitly described in the text
No - Not described in the text

EM5: 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

EM6: 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

EM7: 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

EM8: 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

EM9: Do they discuss and provide reasoning for "why" the results had the given outcome?

Yes - There is a discussion of why a particular outcome occurred in the study. Rather than presenting only the results, the authors explain "why" such results were obtained.
No - No reasoning for the outcome of the study is given.

EM10: Is there a dedicated discussion of the threats to validity to the experiment (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 Proof

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

3.3 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
- 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, Amiangshu Bosu, Christopher S. Corley, Michael Felderer, Matthias Gander, Jason King, Sedef Kocak, and Joumi Markkula, Markku Oivo Clemens Sauerwein, and Laurie Williams.