# Characterizing Scientific Reporting in Security Literature: An analysis of ACM CCS and IEEE S&P Papers

Morgan Burcham
University of Alabama
mburcham@crimson.ua.edu

Mahran Al-Zyoud
University of Alabama
mmalzyoud@crimson.ua.edu

Jeffrey C. Carver
University of Alabama
carver@cs.ua.edu

Mohammed Alsaleh
UNC at Charlotte
malsaleh@uncc.edu

Hongying Du
NC State University
hdu2@ncsu.edu

Fida Gilani
UNC at Charlotte
sgillan4@uncc.edu

Jun Jiang
UNC at Chapel Hill
jiangcj@cs.unc.edu

Akond Rahman
NC State University
aarahman@ncsu.edu

Özgür Kafalı
NC State University
rkafali@ncsu.edu

Ehab Al-Shaer
UNC at Charlotte
ealshaer@uncc.edu

Laurie Williams
NC State University
williams@csc.ncsu.edu

## ABSTRACT

Scientific advancement is fueled by solid fundamental research, followed by replication, meta-analysis, and theory building. To support such advancement, researchers and government agencies have been working towards a "science of security". As in other sciences, security science requires high-quality fundamental research addressing important problems and reporting approaches that capture the information necessary for replication, meta-analysis, and theory building. The goal of this paper is to aid security researchers in establishing a baseline of the state of scientific reporting in security through an analysis of indicators of scientific research as reported in top security conferences, specifically the 2015 ACM CCS and 2016 IEEE S&P proceedings. To conduct this analysis, we employed a series of rubrics to analyze the completeness of information reported in papers relative to the type of evaluation used (e.g. empirical study, proof, discussion). Our findings indicated some important information is often missing from papers, including explicit documentation of research objectives and the threats to validity. Our findings show a relatively small number of replications reported in the literature. We hope that this initial analysis will serve as a baseline against which we can measure the advancement of the science of security.

## Keywords

Science of Security, Literature Review

## 1. INTRODUCTION

Sustained scientific advancement in a field of study requires a significant level of effort from disparate members of a community. Included in these efforts are both high-quality foundational work by community members and viable methods of communicating that work to the larger community. The communication phase is important in that it allows community members to review, understand, analyze, question, replicate, and extend the published results to deepen and expand the overall knowledge of the community and the ability of research results to impact practice.

A key tenet of scientific investigation is the identification and understanding of the fundamental relationships among variables that contribute to or determine the results observed in individual studies. Often researchers cannot identify or understand these relationships on the basis of an individual study. A research community must be able to examine the results and important causal factors from multiple related studies to identify patterns that can provide the deeper insight needed to make progress in the foundational scientific understanding of a field.

However, the practice of cybersecurity today is frequently reactive rather than proactive. That is, because of the frequency and severity of constantly looming threats, organizations often operate in a mode of reacting to attacks after they occur by patching individual vulnerabilities that provided the opening for the attack. For the community to advance from reactive to proactive solutions, we need to gain a better understanding of scientifically-based design principles that allow us to build security in from the beginning. Such an approach would provide more defense against broader classes of both known and unknown attacks.

Recognizing this need, government agencies and security researchers have begun working towards a "science of security". To facilitate additional advances in the scientific underpinnings of security, the research community needs to be able to perform scientific tasks like replication, meta-analysis, and theory building. As indicated in the JASON

*HoTSoS'17, April 3-5, Hanover, Maryland*
© 2017 ACM. ISBN 978-1-4503-5274-1…$15.00
DOI:10.1145/3055305.3055307

report, “The highest priority should be assigned to establishing research protocols to enable reproducible experiments” [3].

The JASON report [3] contends, “There is every reason to believe that the traditional domains of experimental and theoretical inquiry apply to the study of cyber-security.” The advancement of science in these traditional domains involves two key requirements. First, members of a community need to be conducting high-quality research addressing relevant problems. Second, the reports describing this high-quality research need to contain the information necessary to allow for replication, meta-analysis, and theory building.

Therefore, the goal of this paper is to *aid security researchers in establishing a baseline of the state of scientific reporting in security through an analysis of the content of papers in top security conferences.* We emphasize that this only on characterizes the completeness of the information in the papers and does *not* judge the quality of the underlying work (which we assume to be of high quality). In our initial work characterizing the proceedings of the *2015 IEEE Security & Privacy* conference [8], we defined a set of rubrics based on literature on scientific evaluation [16, 17, 18, 19, 23]. The current paper expands on that original paper by answering two questions about the proceedings of the *2015 ACM CCS* and *2016 IEEE Security & Privacy* conferences:

**RQ1:** *What are the characteristics of the artifacts and evaluations contained in the papers?*

- **RQ1.1:** *What types of artifacts are evaluated (i.e., models, languages, protocols, processes, tools, or theories)?*
- **RQ1.2:** *What methods are used for artifact evaluation (i.e. empirical study, proof, or discussion)?*
- **RQ1.3:** *Do papers build on or extend prior work?*
- **RQ1.4:** *Are there trends in the relationship between artifact type and evaluation method?*

**RQ2:** *Do the papers contain all the information necessary to support the science of security?*

The analysis in this paper will help establish a baseline against which to measure progress related to the science of security. We also hope that this paper can serve as an encouragement to members of the community regarding which information should be reported in papers to support the overall advancement of the field.

The remainder of this paper is organized as follows. Section 2 describes background information. Section 3 discusses related work. Section 4 defines the rubric used for paper analysis. Section 5 explains the methodology we used to analyze the papers. Section 6 contains the results of the analysis. Section 7 provides some overall observations across the whole set of papers. Section 8 describes our lessons learned while conducting this study. Section 9 enumerates the threats to validity. Section 10 summarizes the paper.

## 2. BACKGROUND

This section provides background information on the key concepts that are important for the advancement of science. Then it discusses some examples of guidelines that support the application and reporting of research.

### 2.1 Replications

A key tenet of science is reproducibility of results. Reproducibility consists of: (1) obtaining the same results of the original study using the same method in the same environment (where possible, i.e. through a virtual machine replicating an environment); and and (2) providing enough information about the study conditions to allow colleagues to build on results and advance scientific progress [9]. Replications expand this definition by attempting to re-execute studies in different environments (driven by conscious changes to increase the robustness of the overall finding). The ability to reproduce results in various contexts allows researchers to evaluate external validity by determining the extent to which the causal relationships and results/findings of the original study can be generalized [7].

In theory, if there is enough detail about the original study, the results can be validated independently by other researchers [26]. Conversely, if this information is lacking, then research findings likely will be isolated to the original paper and scientific progress will be slower. Therefore, it is important for researchers to provide the right information in their research reports to reduce the overhead introduced when potential replicators have to solicit information from the original researchers.

The problem of replication has been discussed and addressed in different ways:

- Medical researchers have been debating on the validity of their published results for some time [1, 20, 15]. A July 2015 article in MedlinePlus [4] reported that researchers could not reproduce half of the 100 publications in premier psychology journals.
- The ACM SIGMOD community awards the Reproducible Label to database papers, which means “The experimental results of the paper were reproduced by the committee and were found to support the central results of the paper. The experiments (data, code, scripts) are made available to the community”[1].
- The Computational Science community has long recognized the need for reproducibility [24, 25, 2], but has yet to develop a comprehensive solution. Recently *ACM Transactions on Mathematical Software* introduced the Replicated Computational Results designation, awarded to papers for which the editors can obtain “independent confirmation that the results contained in the manuscript are correct and replicated” [14].

### 2.2 Meta-Analysis

Meta-analysis is a systematic approach to analyze the results of a set of previously conducted research studies to derive conclusions about the entire body of research [12]. This process requires researchers to follow a clearly defined process to identify all relevant studies in the literature. Based on those identified studies, meta-analysis uses various statistical approaches to determine the existence, size, and variability of an overall effect. A meta-analysis helps researchers answer new questions, resolve conflicting results, and generate new hypotheses [12]. The abundance of studies and clinical trials on various treatment protocols has provided the necessary data for medical researchers to frequently and successfully apply meta-analysis to draw general conclusions from the disparate studies [29, 11].

Similarly, as the body of studies in the security domain grows, meta-analysis will become an increasingly important

[1] http://db-reproducibility.seas.harvard.edu

method for drawing overall conclusions that can guide future research and practice. To successfully enable the use of meta-analysis in security research, it is essential that researchers provide thorough documentation of their studies.

## 2.3 Theory Building

One goal of research is to build the knowledge required to organize findings into coherent statements about the domain. A theory is the belief that there is a pattern in related observations [10]. A theory can help to fill in the gaps in current knowledge. Further, a scientific theory is an explanation of some phenomenon that is acquired via the scientific method and confirmed through repeated observation and experimentation[2]. Therefore, researchers can test theories and use them to make falsifiable predictions [21].

In security science, as an applied science, theories are important to guide people when making choices about the application of existing solutions to unknown problems. As a constantly evolving field, security science requires a continual growth of the body of knowledge and a deeper understanding of the underlying theories. This knowledge will allow researchers to communicate solutions to practitioners and develop common research agendas.

## 2.4 Guidelines for Reporting Research

As a community coalesces around accepted study design and result reporting mechanisms, it becomes easier for community members to follow appropriate methods [31]. Clear guidelines help a field mature over time, e.g. medicine [5, 27], psychology [28, 13], and social science [6, 22]. Increased maturity in these fields makes it easier to perform tasks like replication, meta-analysis, and theory building.

Analysis of literature from these fields provides insight into balancing scientific rigor and practical relevance. To have the most impact, a community must understand how to report studies (both the designs and the results), how to describe design alternatives, and how to interpret the results for practical benefit. One prime example of such a community is the Cochrane Collaboration in medicine (http://www.cochrane.org). The stated goal is to provide a world of "improved health where decisions about health and health care are informed by high-quality, relevant, and up-to-date synthesized research evidence." To achieve this goal, the community follows a set of principles that ensure: collaboration, avoiding duplicated effort, minimization of bias, relevance, quality assurance, and wide participation. Because members of the community understand how their research results will be used to further larger goals, there is an understood approach to study reporting. As a result, the Cochrane Collaboration has been able to analyze disparate research results to produce reports that transform the way health decisions are made. While we do not necessarily advocate this exact model for the security research community, the benefits that can be seen from rigorous study reporting should be informative.

# 3. RELATED WORK

This section describes related work about the Science of Security and previous literature reviews.

## 3.1 Science of Security

In 2010, JASON was tasked by the US Department of Defense to perform a study on the interplay of science and cybersecurity. The resulting report indicated that a most important attribute is "the construction of a common language and a set of basic concepts about which the security community can develop an understanding." [3]

This work is part of the U.S. National Security Agency *Science of Security Lablets*[3] which seek to develop the scientific underpinnings of security and build a body of knowledge to support rigorous design methodologies. Other similar security research programs around the world include:

- The Team for Research in Ubiquitous Secure Technology (TRUST) is a US National Science Foundation Science and Technology Center based out of the University of California at Berkeley, with the goal of developing "cyber security science and technology that will radically transform the ability of organizations to design, build, and operate trustworthy information systems for the nation's critical infrastructure"[4];
- The MURI project sponsored by the US Air Force OSR based out of Carnegie Mellon University, Cornell University, Stanford University, the University of California at Berkeley, and the University of Pennsylvania, with the goal of "advancing a science base for trustworthiness by developing concepts, relationships, and laws with predictive value"[5]; and
- The Research Institute in Science of Cyber Security based out of the University College of London, with the goal of "giving organizations more evidence, to allow them to make better decisions, aiding to the development of cybersecurity as a science"[6].

## 3.2 Related Studies

We previously analyzed 55 papers of the 2015 *IEEE Symposium on Security and Privacy* with a focus on the completeness of the information provided about the evaluation methods [8]. We used a set of rubrics to determine the type(s) of artifacts being evaluated, the evaluation method, and the completeness of the details for that evaluation method. Some key observations from this study include: (1) tools and processes were the most commonly evaluated artifacts; (2) most papers did not compare their results against a baseline; (3) many papers lacked a clear description of research objectives; and (4) most papers did not discuss threats to validity or study limitations. Section 7.3 compares the results of our current paper with these prior results.

The Asymmetric Resilient Cybersecurity Initiative at Pacific Northwest National Laboratory[7] developed a Science Council [30]. This organization provides insights on applying the scientific method in cybersecurity research and describes the initial impacts of applying the science practices to cybersecurity research and identified eight practices as beneficial in improving the quality of experiments and generating repeatable outcomes: Defining a Tractable Problem, Preliminary Data Assessment, Developing Falsifiable

[2] Based on a definition provided by the National Academy of Sciences (http://www.nap.edu/read/6024/chapter/2#2)

[3] http://cps-vo.org/node/5253

[4] https://www.truststc.org/about/

[5] https://sites.google.com/site/sosmuri/

[6] http://www.riscs.org.uk

[7] http://cybersecurity.pnnl.gov/arc.stm