11

Overcoming Barriers to Empirical Cyber Research

Anne E. Boustead and Scott J. Shackelford

1 INTRODUCTION

Empirical studies have the potential to both inform and transform cyber peace research. Empirical research can shed light on opaque phenomena, summarize and synthesize diverse stakeholder perspectives, and allow causal inferences about the impact of policymaking efforts. However, researchers embarking on empirical projects in the area of cyber peace generally, and cybersecurity specifically, face significant challenges – particularly related to data collection. In this chapter, we identify some of the key impediments to empirical cyber research, and suggest how researchers and other interested stakeholders can overcome these barriers. While these issues stretch across different categories of research designs, some barriers are likely to generate more concern in the contexts of certain types of research questions, as is summarized in Table 11.1. Furthermore, while these obstacles are by no means unique to empirical cyber research, they are particularly salient in this context – and we focus on mechanisms for addressing these barriers that are most likely to be useful to cyber researchers.

2 BARRIERS TO EMPIRICAL CYBER RESEARCH

2.1 *Cyber Decisions and Outcomes Are Difficult to Observe*

Difficult-to-obtain data are a common and persistent problem for empirical cyber researchers. Although there are some publicly available data on cyber policies and outcomes (Federal Bureau of Investigation, 2020; Indiana Attorney General, 2020; National Conference of State Legislatures, 2020), these datasets can be fragmentary, and are few and far between. Data that have become available through less traditional means – such as the leaking of information after a data breach – can provide crucial insights into important, previously unobservable phenomenon, but their use in research raises novel and difficult ethical questions (Boustead & Herr, 2020). In the absence of publicly available datasets, researchers conducting empirical cyber

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press

TABLE 11.1 *Most salient barriers to addressing different types of empirical cyber research questions*

Type	Description of Type	Example Cyber Question	Most Salient Barrier
Exploratory	Focus is on describing and explaining phenomena; may be used to analyze the range of variation in a phenomenon	How do organizations decide whether to use an external cyber risk decision-making framework?	*Empirical cyber research projects frequently require expertise from multiple domains,* complicating systematic exploration of cyber phenomena
Parameter Estimation	Focus is on quantitatively estimating characteristics of a population in a statistically valid way; generally requires particular kinds of random sampling	How many hours of cybersecurity training do hospital employees receive every year?	*Research may only be possible in a narrow range of contexts,* making it difficult to systematically observe a population of interest
Causal	Focus is on establishing whether a cause-and-effect relationship exists between two characteristics of a phenomenon	Do policies requiring regular password changes reduce the frequency of successful cyberattacks?	*Cyber decisions and outcomes are difficult to observe,* making it difficult to identify and evaluate policymaking

projects must rely more heavily on data collection, increasing the time, effort, and resources necessary to conduct research.

Data collection in empirical cyber research is further complicated by the range of actors involved in cyber policy, and differences in how these actors document and disclose their cyber decision-making. Government cyber policymaking is typically memorialized in publicly released documents – including statutes and judicial opinions – which can be analyzed and used to evaluate the effects of these policies on important outcomes (Romanosky, Telang, & Acquisti, 2011). However, much cyber policymaking occurs on an organizational level through decisions made by specific companies and groups about how to manage their own cyber practices (Harknett & Stever, 2009). This decision-making frequently does not result in public documentation, and organizations may be highly reticent to disclose details of their cyber practices due to concerns about security, brand reputation, or liability.

Researchers cannot evaluate policies that they cannot observe and, perhaps more insidiously, efforts to evaluate observable government policies may be undermined by simultaneous and unobservable organizational decision-making. For example, a heavily publicized data breach event could result in observable legislation mandating

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press

employee cybersecurity training, as well as unobservable changes in corporate cyber infrastructure. If the frequency of data breaches declines after the legislation becomes effective, researchers may attribute this change to the legislation without being aware of the confounding and unobservable changes in corporate cyber infrastructure. Reluctance to provide information about cyber decision-making can also result in low survey response rates, making it difficult to accurately estimate how often organizations are adopting particular cyber practices.

2.2 *Empirical Cyber Research Projects Frequently Require Expertise from Multiple Domains*

Cyber systems consist of more than just technology; they also include the people and organizations involved in using and managing cyber systems. Consequently, empirical cyber research often requires data and analytic techniques from multiple domains and disciplines. For example, a project studying how the passage of data breach notification laws impacts cybersecurity behaviors and outcomes would require expertise in law, behavioral sciences, and computer science (Murciano-Goroff, 2019). The range of expertise necessary to conduct these projects generally suggests the need for an interdisciplinary research team. However, differences in the expectations and incentives placed upon researchers in different disciplines may make collaboration difficult.

2.3 *Research May Only Be Possible in a Narrow Range of Contexts*

While some categories of research questions can be answered with only a limited range of observations, others require either a broader scope of data collection or the use of specialized sampling techniques. This is particularly important when trying to describe a characteristic of a population; for example, when estimating the percentage of Fortune 500 companies that employ a Chief Information Security Officer, or how many hours of cybersecurity training hospital employees receive every year. In order to estimate these characteristics in a statistically valid way, researchers must be able to select individuals from the population to observe so that (1) every member of the population could potentially be studied, and (2) the researcher knows how likely it is that each member would be selected. This process – which is known as conducting a probability sample – generally requires identifying every member of the population and selecting members at random to observe (Groves et al., 2011). In the case of cyber peace research, identifying every member of the population can be particularly difficult, especially when researchers are trying to estimate characteristics of technical populations (such as malware) rather than human ones. Even when it is possible to address a research question by studying a narrower population, this choice may impact the generalizability of the research (Lee & Baskerville, 2003). As a result, both researchers and policymakers must be careful when trying to

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press

generalize the results of the study. For example, further research would be needed to determine whether the results of a survey of cybersecurity practices conducted in Indiana could be generalized to other states (Boustead & Shackelford, 2020).

3 OVERCOMING BARRIERS

Although these barriers pose significant challenges to empirical cyber research, they are not insuperable. In the remainder of this document, we identify several practices that individual researchers, universities, and other organizations could adopt to facilitate empirical cyber research.

3.1 *Incentivize Interdisciplinary Research Teams*

To overcome these difficulties, exploratory cyber research projects may especially benefit from an interdisciplinary team, with expertise in technology, policy, law, and behavioral science. Fortunately, there is a long history of interdisciplinary collaboration in cyber research, including cross-disciplinary conferences, journals, academic programs, and other initiatives. In order to further encourage interdisciplinary cyber research, we would suggest that academic leaders in multiple disciplines make clear how interdisciplinary research will be accounted for during the tenure and promotion process (Benson et al., 2016). Additionally, researchers across multiple disciplines should be encouraged to engage in cross-disciplinary teaching experiences in order to educate future researchers and decision-makers to engage in interdisciplinary research, and create partnerships between disciplines to facilitate future research. An example of this approach in action is the IU Cybersecurity Clinic, which is unique in both its interdisciplinary breadth, as well as the fact that it is open to all graduate students across campus and offers applied service-learning opportunities to assist local and state-level critical infrastructure providers.

3.2 *Partnerships Are Key*

Oftentimes, empirical cyber research questions may be of interest to a variety of stakeholders in the public and private sectors. A state government may be interested in information about the uptake of cybersecurity practices amongst businesses in their jurisdiction, while a trade group might be interested in perceptions of privacy protections amongst their constituents. For example, the authors of this paper have collaborated with the State of Indiana to field a survey on cybersecurity practices amongst organizations in Indiana in order to address both academic and policy questions on cybersecurity decision-making. Under these circumstances, partnering with stakeholders has the potential to facilitate and improve cyber research. Research partners can provide insights into the phenomena in which they are involved, and insider knowledge about how policies are implemented in practice

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press

can provide a critical counterpoint to academic expertise. Furthermore, stakeholders are often experts in their own decision-making, and emic explanations about their policies and practices can be irreplaceable.

Research partnerships with public or private stakeholders can take on a number of forms. Researchers can consult with stakeholders during project development in order to identify potential causal mechanisms, locate existing data sources, and preview interview questions to determine whether they are likely to elicit relevant information. Stakeholder research partners may also be willing to facilitate data collection by distributing surveys or providing introductions to potential interview subjects. Because they are likely interested in the results of research, stakeholder partners may also be helpful in disseminating the results of research projects and encouraging consideration of policy recommendations resulting from the project.

While partnerships with public or private organizations can greatly benefit empirical cyber projects, researchers must be mindful of several potential complications. Public and private organizations may have a more limited remit than the population that might be of interest to the researcher. For example, a state or local government may be able to provide data about their own jurisdiction, and an industry trade group may be able to assist in distributing a survey to their members. These constraints can generally be addressed by narrowing the research question to focus on the population for which data are available; however, a more limited study may be less generalizable, and efforts to use these studies for policy decision-making in other areas must account for differences in context. It may also be helpful to repeat research in multiple contexts in order to explore the circumstances under which the results of the study hold.

Researchers who partner with public or private entities should also be prepared to navigate potential conflicts between the goals of the research partner and the goals of academic research. Organizations may partner with researchers because they have an interest in obtaining answers to particular questions or learning more about phenomena that affect them. Researchers may consider expanding the scope of their research to ensure that questions of interest to the partner are also addressed, and seeking out partnerships where there is a natural overlap in the questions of interest. However, partners should never have control or veto power over whether the results of the research are released. In order to ensure that partnering organizations can benefit and learn from the research, researchers should consider ensuring that results are available in formats that are usable by the partner; for example, publishing reports and podcasts, as well as journal articles.

3.3 *Publish Cyber-Related Data*

The field of empirical cyber research as a whole would benefit tremendously from an increase in the scope of publicly available data on cyber policies, decisions, and outcomes. Publicly available data facilitate and incentivize research by lowering the

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press

costs of undertaking projects. They also create efficiencies by ensuring that data collected are available to many researchers, reduce the burden on participants who may be asked to participate in multiple studies unless data collection is coordinated, and increase transparency in both research and policymaking (Napoli & Karaganis, 2010). Organizing the release of datasets could also serve as a mechanism for promoting high-quality cyber research if the data released are valid and reliable.

There are a number of mechanisms for ensuring the availability of empirical data on cyber phenomena. Over the short term, the publication of an annotated bibliography describing the datasets that are available, and highlighting where the collection of data in other domains has touched upon cyber-related issues, would both make those data more accessible to researchers and identify gaps in current data availability. Efforts could then be undertaken to expand current data collection projects to include information about cyber-related issues where relevant; for example, adding a question to a survey of hospital administrators to ask about their cybersecurity practices. Finally, surveys and other data collection projects focused on cyber issues could be undertaken and expanded, with priority given to efforts that can be repeated on a yearly basis in order to observe changes over time. These efforts could be facilitated through collaboration with existing public–private cyber partnerships, such as Executive Councils on Cybersecurity and organizations designed to share cyber threat information within sectors, such as information sharing and analysis centers and information sharing and analysis organizations. There is no one-size-fits-all model, but through experimentation and deeper partnerships, we may glean a more accurate picture of the cyber peace landscape.

REFERENCES

Benson, M. H., Lippitt, C. D., Morrison, R., Cosens, B., Boll, J., Chaffin, B. C., … Link, T. E. (2016). Five ways to support interdisciplinary work before tenure. *Journal of Environmental Studies and Sciences*, 6(2), 260–267.

Boustead, A. E., & Herr, T. (2020). Analyzing the ethical implications of research using leaked data. *PS: Political Science & Politics*, 53(3), 505–509.

Boustead, A. E., & Shackelford, S. (2020). State of Hoosier Cybersecurity. Retrieved from www.ibrc.indiana.edu/studies/State-of-Hoosier-Cybersecurity-2020.pdf?_ga=2.164292790.820309617.1609805269-789814884.1603708149

Federal Bureau of Investigation. (2020). 2019 Internet Crime Report. Retrieved from https://pdf.ic3.gov/2019_IC3Report.pdf

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey methodology* (Vol. 561). John Wiley & Sons.

Harknett, R. J., & Stever, J. A. (2009). The cybersecurity triad: Government, private sector partners, and the engaged cybersecurity citizen. *Journal of Homeland Security and Emergency Management*, 6(1), 1–14.

Indiana Attorney General. (2020). Identity Theft Prevention. Retrieved from www.in.gov/attorneygeneral/2874.htm

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research*, 14(3), 221–243.

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press

Murciano-Goroff, R. (2019). Do Data Breach Disclosure Laws Increase Firms' Investment in Securing Their Digital Infrastructure? Paper presented at the Workshop on the Economics of Information Security.

Napoli, P. M., & Karaganis, J. (2010). On making public policy with publicly available data: The case of US communications policymaking. *Government Information Quarterly*, 27(4), 384–391.

National Conference of State Legislatures. (2020). Security Breach Notification Laws. Retrieved from www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

Romanosky, S., Telang, R., & Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2), 256–286.

https://doi.org/10.1017/9781108954341.011 Published online by Cambridge University Press