Online Appendix – A Tale of Two Cybers

Lennart Maschmeyer, Ronald J. Deibert & Jon R. Lindsay

This online appendix contains the following three parts. Part A1 offers definitions of key terms. A2 provides further background on the influence of commercial threat reporting on academia and public policy, an alternate interpretation of the role of vendors and a discussion of the underlying causes of the selection bias in reporting we expect to lead to a distorted picture of cyber conflict. A3 covers further details of the research design, namely reliability tests, data collection challenges, the full codebook and a list of all coded reports.

A1: Definitions

Threat actor is a common term in commercial threat reports for the actors behind cyber operations or incidents.

Advanced Persistent Threat (APT) is a term introduced by the firm Mandiant to denote the most advanced threat actors, which are defined by their persistence and sophistication (Mandiant 2013). It has since become commonly used within the industry, academia as well as the policy world.

Tactics, Techniques and Procedures (TTP) refer to the methods and tools used by a threat actor, and has become an important means of distinguishing and identifying different threat actors. The National Institute of Standards and Technology accordingly simply defines TTP as "the behavior of an actor" before expanding that "a tactic is the highest-level description of this behavior, while techniques give a more detailed description of behavior in the context of a tactic, and procedures an even lower-level, highly detailed description in the context of a technique" (NIST 2020)

Civil Society Organizations (CSO) are defined as independent organizations and individuals who engage in nonviolent political activity, including activists, dissidents, journalists, academics and nongovernment organizations. This is a heterogenous set of actors, but with one distinguishing characteristic: political engagement without being part of government, and without the use of coercion. Accordingly, Mary Kaldor highlights the emphasis on rule-based behavior rather than coercion (Kaldor 2013, 1). Importantly, since the distinguishing characteristic of civil society is non-violent political engagement, we only include journalists and academics if this condition applies to them.

Targeted threats refer to cyber operations aimed at specific targets that attempt to gain access and compromise systems persistently and over long periods of time, often as part of campaigns (Hardy et al. 2014, 527). They often employ customized TTP that are carefully calibrated to the vulnerabilities of the chosen targets.

A2: What do We Know?

A2.1 The Influence of Commercial Threat Reporting on Academic and Policy

The most dramatic example of the policy relevance of threat reporting is the vendor CrowdStrike's role in the ongoing controversy surrounding interference in the 2016 Presidential Elections by Russian threat actors. Not only was the Russian intrusion into the network of the DNC first revealed in a commercial report (CrowdStrike 2016), but the vendor behind this report has since become implicated in a creative conspiracy theory promoted by US President Donald Trump. This theory claims the DNC server was secretly moved to Ukraine, where it is held by CrowdStrike, based on the false claim that the company's founder Dmitri Alperovitch, is a Ukrainian national (CNN 2019). Significantly, President Trump asked Ukrainian President Zelinskiy for information on the server, as part of the alleged quid-pro-quo request that formed the basis for the impeachment process against the President. CrowdStrike has since updated its blog post on the DNC intrusion to rebuke such conspiracy theories, and underline their non-partisan role as a business (CrowdStrike 2020). Other examples include the GRIZZLY STEPPE report by the US Department of Homeland Security that outlines activities and methods of Russian threat actors. The report "encourages network defenders, threat analysts, and general audiences to review publicly available information to develop a better understanding of the tactics, techniques, and procedures (TTPs) of APT28 and APT29" and prominently cites a list of commercial threat reporting (US CERT 2017, 2).

The latter example highlights how commercial threat reporting influences government's own threat assessments. The GRIZZLY STEPPE report mentioned above is part of a nascent initiative by the United States Department of Homeland Security to provide public report, an effort mirrored by the cybersecurity agencies in the UK and Germany. In both cases, these government-issues reports heavily rely on commercial threat reports, often simply summarizing linked commercial reports. For example, the United Kingdom's National Cyber Security Centre's weekly threat report from January 10 has two main news items, both of which summarize research by two commercial firms. Similarly, the German Intelligence Service's 'Cyber Brief No. 2' from 2018 leverages Kaspersky research to highlight ongoing phishing campaigns targeting German entities.

For an academic example, Ben Buchanan's influential book on the cybersecurity dilemma heavily relies on data from commercial reporting in its empirical analysis. As Buchanan underlines, "the primary sources for these [empirical] sections are often professional reports on noteworthy intrusions written by computer forensic analysts" (Buchanan 2017, 10). Moreover, he stresses that "these sources prove to be of great value in understanding the known cyber operations of states and in presenting important cases" and that "one cannot credibly claim to understand states' current approaches to cyber operations without examining this detailed evidence of actual behavior" (Buchanan 2017, 10). While the data from commercial reporting is definitely useful, how representative they are of 'current approaches to cyber operations' is precisely our contestation. What is reported shapes not only perception among academics, but also

policy-makers. Hence, it is legitimate to consider to what extent the reports Buchanan holds to provide data on 'current approaches to cyber operations' are not only sources of data, but instead causes of the "perception of offense-dominance" that exacerbates the looming cybersecurity dilemma and which "seems particularly pronounced among policy-makers" (Buchanan 2017, 109).

Similarly, Joseph Nye's recent article on deterrence reflects prevailing wisdom informed by commercial threat reports in describing the cyber threat landscape as follows: "there is a wide range of cyber threats, including war, espionage, sabotage, and disruption" (Nye 2017, 47), and identifies possible means of deterring nation-state attacks on critical infrastructure (Nye 2017, 56–61). While Nye does not directly cite commercial reporting, in building this argument he mostly relies on newspaper and magazine articles that prominently feature representatives from private vendors and their reporting (pp. 44-48). For example, one of the key sources underlining threats to critical infrastructure is a New York Times article by Nicole Perlroth, whose key conclusions are provided by industry representative.[1] This evidence-based threat perception is far more nuanced than early cyberwar prophecies, but it reflects priorities in commercial reporting and neglects one key target category: CSOs. To be sure, Nye's primary focus is on interstate conflict, yet he does explicitly include non-state actors and criminal activity in his threat model (Nye 2017, 69), underlining the bracketing of civil society.

A2.2 Commercial Threat Reporting as a quasi-academic Enterprise

A less cynical interpretation of commercial reporting is that it not only serves business interests, but plays a quasi-academic role enabling cybersecurity researchers to communicate findings benefiting the broader security community. By communicating findings publicly, firms open themselves up to criticism and ridicule from competitors if analysis is flawed. Hence, leaders have reputational incentives to perform better analyses. Open peer-to-peer sharing also accords with the norms of hacker communities and firms aiming to retain talented reverse engineers are incentivized to let them participate in such networks. Unfortunately, these non-advertising incentives also tend to bias reporting toward the most technically sophisticated operations against the most high-profile targets. Moreover, they are all complementary with the overriding marketing function of public reporting. Hence, even if this alternate interpretation is accurate, we would still expect it to lead to the same systematic selection bias we predict.

A2.3 High profile

Profile in this context is not a scientific measure, but in general an indicator of public attention. Accordingly, the Cambridge English Dictionary defines it as "attracting a lot of attention and interest from the public and newspapers, television etc." (Cambridge English Dictionary n.d.). Key indicators of a high-profile actor as perceived by the target audience are public and media

[1] The report cites "Joe Weiss, a crusader for industrial control security and founder of Applied Control Solutions, a consulting firm" who paints a menacing picture in suggesting that "not only do we not have the mitigation, we don't even have any type of adequate forensics to know this [critical infrastructure sabotage] is happening" (Perlroth 2015).

attention. From the perspective of threat intelligence vendors, the highest profile targets and threat actors are those maximizing attention of the intended audience to the report in question.

A2.4 Expected Characteristics of Targeted Threats against Civil Society

First, operations targeting civil society tend to be technically unsophisticated, employing generic tools and known vulnerabilities. Since civil society actors typically lack resources and expertise necessary for strong cybersecurity, cyber operations targeting them can often succeed with generic and unsophisticated methods (Crete-Nishihata et al. 2014, 3). Commercial spyware forms an exception, yet is unlikely to be reported on due to legal and political risks involved. Commercial 'spyware' offers highly sophisticated means of exploitation and is evidently used by governments for surveillance of civil society, as the Abdulaziz case highlights. Considering their sophistication, one could expect commercial reporting to prominently cover spyware operations. However, spyware vendors typically restrict usage to 'lawful intercept', rendering cases involving international espionage or intellectual property theft unlikely. Moreover, reporting on breaches of such usage restrictions carries significant legal risks since vendors may threaten legal action—as Citizen Lab has experienced (R. Deibert 2016). Considering these disincentives, commercial reporting is unlikely to cover spyware prominently.

Second, their lack of resources renders civil society actors unattractive potential clients, and thus a low priority in reporting. Some civil society actors receive a high level of public and media attention as a result of their activism, however, which may promise enough potential attention to threat reporting on this actor to offset above disincentives in individual cases.

Third, campaigns targeting civil society are regularly pursued by the same actors as those targeting governments, militaries and large corporations. The Citizen Lab's Communities @ Risk study (2014) showed that half of the campaigns targeting CSOs it tracked over a course of four years had "connections to threat actors, previously reported to have targeted government and private industries." (Crete-Nishihata et al. 2014, 2). We may thus expect these types of campaigns to be prominently reported by private vendors. Yet trends towards the use of generic TTP identified in the same study may offset these incentives—especially if targeting focuses on poorer regions, i.e. the Global South.

A3: Hypotheses and Research Design

A3.1 Selection Criteria

For this study, we only included reports on targeted threats. Following our definition, these are reports that publish findings of investigations into cyber operations by specific threat actors that target specific victims, characterized by persistent efforts often covering a longer period of time, and using customized TTP calibrated to the chosen victims. We distinguish between three types

of reports: (1) focused APT reports that examines tradecraft of a specific actor, including political analysis into objectives and possible attribution attempts, (2) surveys that summarize threat activity by one or more APTs over an extended timeframe, and (3) technical reports that focus exclusively on technical aspects of the TTP employed by a known threat actor. Hence, we do include technical reports that discuss specific exploits used by a specific threat actor, but we do not include general reporting on vulnerabilities in known software.

A3.2 Rival Explanations

There are several plausible alternate explanations. First, reported threat patterns may accurately reflect reality, and correlations may be purely coincidental. Second, divergences between commercial and independent reporting may reflect independent reporting's own bias(es), or other differences among these two slices of data. Third, the lack of reporting on specific regions may reflect lack of knowledge rather than under-prioritization. Commercial vendors rely on telemetry data from their security products to identify threats and may lack visibility into specific regions or sectors. However, a lack of telemetry data itself likely reflects underlying business interests: telemetry data is drawn from existing customers, and if there are no customers in a region, that region has not been developed. In short, this analysis allows us to verify whether reporting patterns correspond to those one would expect if reporting was indeed biased across the hypothesized selection criteria.

A3.3 Causal Mechanism and Limitations of the Data

Because the selection and decision-making processes of threat reporting happen behind closed doors, however, there is insufficient data to prove the causal relationship between selection criteria and reporting bias we hypothesize. To address this issue, our initial research design envisions a set of structured interviews with researchers and management at a sample of threat intelligence firms to trace this decision-making process. We designed detailed questionnaires and robust anonymization procedures approved by our University's Research Ethics Board. However, multiple attempts at recruiting interview subjects remained fruitless and ultimately only one researcher was willing to be interviewed. Structured interviews were a core part of our original research design, but despite multiple recruitment attempts at multiple firms, only one researcher was willing to be interviewed anonymously. As in any competitive industry, these firms carefully guard their internal processes and organizational structures—which is understandable, but increases the opaqueness of threat reporting as a data source. This lack of responses itself constitutes a potential data point on the low prioritization of civil society threats. In any case, there is insufficient data to conclusively establish the causal mechanisms in the decision-making process. Therefore, we examine whether the *outcome* of this process conforms to the predictions of our theory.

A3.4 Codebook and Reliability Tests

Below is an overview of our coding criteria for each variable we tracked in reporting.

Variable Name	*Criteria and Description*
CSO mentioned	This is a binary variable, recording whether a report mentions targeting of civil society organizations (Y) or not (N). Civil Society Organizations (CSO) are defined as independent organizations and individuals who engage in nonviolent political activity, including activists, dissidents, journalists, academics and nongovernment organizations. We exclude think tanks because they are often government-funded and carry out research Importantly, since the distinguishing characteristic of civil society is non-violent political engagement, journalists and academics are only to be counted as civil society actors if this condition applies to them. Importantly, reflecting our focus, we only code the three relevant variables below (CSO Focus, Attribution and CSO country) for those reports that do mention civil society.
CSO Focus	Tracks the focus on CSO targeting among reports that do discuss such targeting. Recorded on a spectrum of 1-3: 1 = Primary focus of the report is on CSO targeting; 2 = Secondary focus on CSO, i.e. the primary focus is on another target, but civil society targeting is discussed in some detail and with some analysis. For example, a report on a specific threat actor may focus on its targeting of government entities, but also discuss targeting of civil society groups by the same actor; 3 = CSO targeting is only mentioned in passing, without any analysis. For example, a report may include CSOs within target lists at the end of the report, but do not discuss this targeting in the report itself. Similarly, a report focusing on a specific threat actor may mention briefly that the same actor has previously targeted civil society, but without further context or analysis.
Attribution	Attribution is reported among a spectrum of confidence, but for simplicity we record it here as a binary value of attribution/non-attribution as well as the name of the state to which an operation is attributed. Attribution is recorded whenever the report in question attributes the activity it discusses to a specific state, regardless of the level of confidence. Attribution is also recorded for reports where the introduction of a threat actor links to previous reporting on the same actor that clearly attributes it to a state. On the other hand, reports in which attribution is merely implied based on circumstantial evidence (i.e. the objectives align with Chinese interests) are not recorded as attribution. Reports making no mention or attempt at attribution are (obviously) also coded as "not attributed".

CSO country	Records the country/countries in which CSOs targeted are located as described by the report. If the report does not distinguish targeted countries by target type, we record all listed countries where victims are located. Country names are recorded in full text, e.g. 'Ethiopia'.
Report Type	This variable is for reference only, and captures the type of report, distinguishing among three different types: 1 = Focused APT report, considers political context and interests, looks into targeting patterns, wider analysis of TTP (beyond purely technical details), attempts attribution; 2 = Threat survey; survey of trends in targeted threats over specific time, or survey of evolution a specific type of threat, or survey of attack trends on a specific sector; 3 = Focus on technical aspects of attack, discussing technical details and indicators while targeting or political context discussed not at all or in passing only
Threat Actor	This category is coded for reference only and records the name of the threat actor to which a cyber operation is attributed, as named by the company publishing the report. If the threat actor is known under other, more common names, this name should be added.

All coding was carried out by a single researcher (the main author) following established practice by specifying clear and unambiguous criteria. The content analyzed is straightforward, 'manifest content'[2], that does not require interpretation of underlying meaning. Consequently, coding involved what Potter and Levine-Donnerstein refer to as 'clerical recording' (Potter and Levine-Donnerstein 1999, 261) where accuracy rather than reproducibility becomes the main determinant of reliability (Potter and Levine-Donnerstein 1999, 271). Hence, once initial coding was completed, we re-coded the entire sample of reports under the microscope (those discussing civil society) to ensure accuracy as well as 'intracoder reliability' (Neuendorf 2002, 268).

To verify the validity of our coding scheme and findings, we then proceeded with an intercoder reliability test.[3] We followed established practice in choosing a random sample consisting of 10% of the reports included in the dataset and calculated both Cohen's Kappa and Krippendorf's Alpha scores for the key variables (Neuendorf 2017, 235; Lombard, Snyder-Duch, and Bracken 2002; Ranganathan 2015). We chose 70 reports based on a random identifier (a randomized number), trained a second researcher in our coding scheme and had the researcher code the entire sample to compare their results with our own. The key variable in this comparison concerns CSO mentions, since this variable is coded for all reports, while additional variables are only coded for the subset of reports that does mention civil society.

[2] Manifest content is defined as "that which is on the surface and easily observable, such as the appearance of a particular word in a written text" (Potter and Levine-Donnerstein 1999, 259)

[3] The coding comparison sheet is available at: https://docs.google.com/spreadsheets/d/1f9tYHclklsakrPFdeUeBNbvwDs5inKzkcJxhVmzAmQM/edit?usp=sharing

Cohen's Kappa Values (Variable: CSO Mentions)				
Agreement	*Expected Agreement*	*Kappa*	*Std. Error*	*Z*
97.14%	75.47%	0.8835	0.1187	7.44

We calculated reliability scores using STATA and its -krippalpha- and -kappa- packages. When comparing the results, we found 97% agreement between coders at an expected agreement of 75%, and a corresponding Cohen's Kappa value of 0.88. The Krippendorf Alpha (2004) score was almost identical, at 0.884 (versus 0.8835 for Cohen's Kappa). These findings strongly support the validity of our coding scheme. While there are no common criteria for minimum scores, a Kappa score of 0.88 reflects 'near-perfect agreement' among coders (Ranganathan 2015, 200).

We observe a similarly high degree of agreement among both coders for the other three variables tracked (CSO focus, Attribution and CSO country). Kappa and Alpha values for these three variables are not significant indicators for reliability due to the small numbers of reports (10 out of the sample of 70 were coded as mentioning civil society by both coders). Yet there were only four individual instances of disagreement across all variables, further underlining the reliability of our analysis. The remainder of this section will briefly discuss these four instances among the subset of reports discussing civil society targeting. There was disagreement among coders whether a report had mentioned civil society or not for two of these reports, reflecting vague language used in the reports in question. Hence, in these two reports only one coder assigned values for the three additional variables.

The third instance of disagreement concerned the level of focus on civil society targeting, which coder 1 coded as primary focus, while coder 2 determined the secondary focus was on civil society. The report on question discusses targeting of the Syrian opposition, which includes armed resistance fighters as well as media activists and humanitarian workers. One coder considered only media activists to be civil society, while the other also included opposition groups in its definition of civil society—even though some of them are armed, the report does highlight that data exfiltrated included minute details of shifting political positions and relationships among opposition groups. Both interpretations are valid according to the coding scheme, and the disagreement results from the lack of distinction between armed and unarmed opposition groups in the report. The fourth instance of disagreement concerned attribution of a campaign to Palestine, where one coder concluded that the report attributes the campaign based on title of the report, "Gaza Cyber Gang" and multiple references throughout the text to the Palestinian origins of the campaign and its alignment with Palestinian interests. The other coder did not agree, because despite these multiple references the report does not draw a linkage to Palestinian government, and does not explicitly discuss attribution. These interpretations are both plausible, although the latter is more closely aligned with the coding book, and again reflect significant ambiguity in the language used in the report in question.

In conclusion, coding results of both coders are closely congruent, indicating only minor disagreements that strongly support the validity of our findings.

A3.5 Link to Data

List of Threat Reports included in the Study

A3.6 H2 Data Collection and Analysis

Few reports specify business sectors and categories used vary across firms. Both commercial and independent reporting do specify the geographical location of victims, however. Hence, if a majority of commercial reporting discusses threats targeting the Global North, it supports our hypothesis. If reporting is evenly distributed between the Global North and South, or if a majority of reporting focuses on the Global South, the hypothesis is invalidated. We then compare geographical distribution between commercial and independent reporting to identify diverging patterns. To ensure a fair comparison, we include only the subsample of commercial reporting discussing civil society. We supplement our dataset with data by AccessNow, an international Internet rights advocacy group that generously provided us with anonymized data from their helpline. This helpline provides a resource for civil society to seek assistance in the event of targeted digital attacks. The data counts the incidents of malware, system compromise and phishing as self-reported by CSOs, categorized by country. We do not consider this data to be more representative than our other sources since it is subject to multiple possible selection biases. Rather, it constitutes an additional slice of data on a type of threat we expect is underreported by commercial actors—and our comparison confirms whether this is the case.

If the geographical distribution of commercial reporting and self-reporting differs along our hypothesized North-South divide, this finding provides further strong support for our hypothesis. If, however, the geographical distribution of self-reporting and commercial reporting are congruent, with independent reporting showing a different distribution, the hypothesis is invalidated—indicating systematic bias in independent reporting instead.

References

Buchanan, Ben. 2017. *The Cybersecurity Dilemma Hacking, Trust and Fear Between Nations*. Oxford: Oxford University Press. http://public.eblib.com/choice/PublicFullRecord.aspx?p=4806712.

Cambridge English Dictionary. n.d. "HIGH-PROFILE | Meaning in the Cambridge English Dictionary." Accessed February 7, 2019. https://dictionary.cambridge.org/dictionary/english/high-profile.

CNN. 2019. "What Is CrowdStrike and Why Is It Part of the Trump Whistleblower Complaint?" *CNN*, September 26, 2019. https://www.cnn.com/2019/09/26/tech/what-is-crowdstrike/index.html.

Crete-Nishihata, Masashi, Jakub Dalek, Ronald Deibert, Seth Hardy, Katharine Kleemola, Irene Poetranto, John Scott-Railton, et al. 2014. "Communities at Risk - Extended Analysis." Citizen Lab. https://targetedthreats.net/media/1-ExecutiveSummary.pdf.

CrowdStrike. 2016. "Bears in the Midst: Intrusion into the Democratic National Committee »." June 15, 2016. https://www.crowdstrike.com/blog/bears-midst-intrusion-democratic-national-committee/.

———. 2020. "Our Work with the DNC: Setting the Record Straight." January 22, 2020. https://www.crowdstrike.com/blog/bears-midst-intrusion-democratic-national-committee/.

Hardy, Seth, Masashi Crete-nishihata, Katharine Kleemola, Adam Senft, Byron Sonne, Greg Wiseman, Phillipa Gill, and Ronald J Deibert. 2014. "Targeted Threat Index : Characterizing and Quantifying Politically-Motivated Targeted Malware." *The 23rd USENIX Security Symposium*, 527–541.

Kaldor, Mary. 2013. *Global Civil Society: An Answer to War*. John Wiley & Sons.

Krippendorff, Klaus. 2004. *Content Analysis: An Introduction to Its Methodology*. SAGE.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2002. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability." *Human Communication Research* 28 (4): 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x.

Mandiant. 2013. "APT1." Mandiant. https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf.

Neuendorf, Kimberly A. 2002. *The Content Analysis Guidebook*. Thousand Oaks, Calif: Sage Publications.

———. 2017. *The Content Analysis Guidebook*. Second edition. Los Angeles: SAGE.

NIST. 2020. "Tactics, Techniques, and Procedures (TTPs)." 2020. https://csrc.nist.gov/glossary/term/Tactics-Techniques-and-Procedures.

Nye, Joseph S. 2017. "Deterrence and Dissuasion in Cyberspace." *International Security* 41 (3): 44–71. https://doi.org/10.1162/ISEC_a_00266.

Perlroth, Nicole. 2015. "Online Attacks on Infrastructure Are Increasing at a Worrying Pace." *Bits Blog* (blog). October 14, 2015. https://bits.blogs.nytimes.com/2015/10/14/online-attacks-on-infrastructure-are-increasing-at-a-worrying-pace/.

Potter, W. James, and Deborah Levine-Donnerstein. 1999. "Rethinking Validity and Reliability in Content Analysis." *Journal of Applied Communication Research* 27 (3): 258–84. https://doi.org/10.1080/00909889909365539.

Ranganathan, Catherine MacPhail, Nomhle Khoza, Laurie Abler, Meghna. 2015. "Process Guidelines for Establishing Intercoder Reliability in Qualitative Studies - Catherine MacPhail, Nomhle Khoza, Laurie Abler, Meghna Ranganathan, 2016." *Qualitative Research*, April. https://journals-sagepub-com.myaccess.library.utoronto.ca/doi/10.1177/1468794115577012.

US CERT. 2017. "Enhanced Analysis of GRIZZLY STEPPE Activity." AR-17-20045. https://www.us-cert.gov/sites/default/files/publications/AR-17-20045_Enhanced_Analysis_of_GRIZZLY_STEPPE_Activity.pdf.