The Role of Cybersecurity in Economic Performance

Estefania Vergara Cobos Selcen Cakir Hagai Mei-Zahav Baran Berkay Barakcin[1]

[**Attribution—**Please cite this work as follows: Vergara Cobos, Estefania; Cakir, Selcen; Mei-Zahav, Hagai; Barakcin Berkay, Baran. 2024. The Role of Cybersecurity in Economic Performance. Washington, DC: World Bank.][2]

Abstract

Utilizing novel data on over 10,000 disclosed cyber incidents across 190 countries and 21 industries, along with data on governments' cybersecurity commitments, this paper investigates the impact of cybersecurity on both industries' performance and macroeconomic outcomes. Specifically, employing an instrumental variable (IV) cross-country, cross-industry model with fixed effects, the study demonstrates that given an exposure level to cyber incidents, industries perform better in countries that have implemented robust national cybersecurity commitments, such as operational legal frameworks, cooperative measures, technical advancements, and capacity-building initiatives.

At the macroeconomic level, the paper reveals a statistically significant negative correlation between the frequency of disclosed cyber incidents and GDP per capita in emerging economies, with more pronounced effects observed in the public sector, information and telecommunications, finance, and education industries.This research pioneers the estimation of the link between cybersecurity and economic performance, addressing the existing uncertainty regarding the returns on cybersecurity investments, particularly in developing countries. The findings underscore the critical importance of cybersecurity as firms navigate the digital era.

Keywords: Cybersecurity, cybercrime, cyber incidents, economic growth, productivity, digital development, digital acceleration, developing countries, cyber resilience, digital economy, digital transformation.

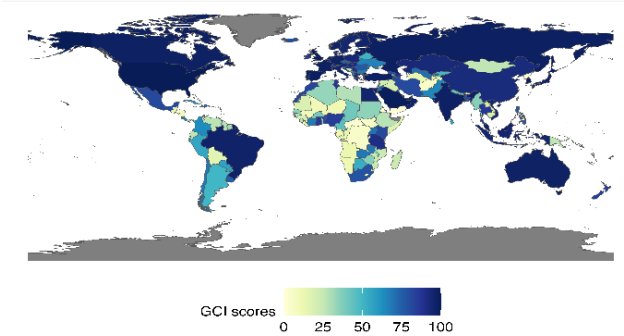
JEL codes: L1, G1, O3, Z01.

[1] This paper is a product of the Chief Economist Office for the Infrastructure Vice-Presidency of the World Bank in collaboration with the Digital Development Global Practice. This research was funded by the World Bank Cybersecurity Multi-Donor Trust Fund as part of a larger effort to study the Economics of Cybersecurity. The authors wish to thank Stephane Straub, Christine Zhenwei Qiang, Casey Torgusson, Bertram Boie, and Anat Lewin for their comments and support, and country donors of the World Bank Cybersecurity Multi-Donor Trust Fund for funding this work. The authors may be contacted at evergaracobos@worldbank.org.

1. Introduction

In early 2023, the U.S. government recognized cybersecurity as an essential element to the basic functioning of the economy by announcing the National Cybersecurity Strategy, which promotes to rebalance the cybersecurity responsibility from individuals, small businesses, and local governments, onto the federal government. Similarly, in 2023, other countries like Japan and Australia pledged to devote more public spending into the defense of the cyberspace through updated national cybersecurity strategies that prioritize protecting businesses and citizens, as well as strengthening critical infrastructure. Thus, in an era of rapid technological development and accelerated digital transformation, cybersecurity is normatively known as a shared responsibility, as the security of a system depends on that of the weakest link. The rapid technological development and the accelerated digital transformation of nations is pushing governments to play active roles in the protection of cyberspace.[3] However, due to the lack of evidence on the effects of cyber incidents on the economy and unclarity in the returns to investments on cybersecurity measures, cybersecurity is still seen as an operational cost rather than a growth or developmental opportunity. This is particularly true amongst developing countries that, given limited budgets and unmet social needs, struggle to prioritize cybersecurity in their investment agendas (see Figure 1).[4]

Figure 1: Cybersecurity commitments by country 2020.



Source: ITU's Global Cybersecurity Index 2020.

Cyber incidents can threaten national and economic security. For example, in 2022, Costa Rica fell victim to a major systemic ransomware attack that compromised the IT systems of almost 30 government agencies including the Ministry of Finance, the Costa Rican Social Security Fund, and the Virtual Tax Administration. This systemic attack lasted for a period of 56 days, in which, for the first time in history, a government declared a state of national emergency due to cyber-attacks, shutting down the computer systems used for collecting taxes, controlling customs, serving beneficiaries, and more. The Costa Rican economy is estimated to have lost about USD 30m per day due to the cyber incident, a total loss of about USD 1.6bn or 2.3% of the country's GDP.

Using novel data on disclosed cyber incidents[5] at national and industry levels and the results of a comprehensive survey to government officials about their cybersecurity commitments, this paper provides first time evidence of the economic effects of cybersecurity.

[3] In early 2023, the U.S. government recognized cybersecurity as an essential element to the basic functioning of the economy by announcing the National Cybersecurity Strategy, which promotes to rebalance the cybersecurity responsibility from individuals, small businesses, and local governments, onto the federal government. Similarly, in 2023, other countries like Japan and Australia pledged to devote more public spending into the defense of the cyberspace through updated national cybersecurity strategies that prioritize protecting businesses and citizens, as well as strengthening critical infrastructure.

[4] According to ITU's Global Cybersecurity Index 2022, a score that reflects countries' commitments in cybersecurity such as technical and legal measures, most low-income countries and low-middle income countries lie in the two bottom quartiles.

[5] A cyber incident is defined as an event or the end result of any single unauthorized effort taken using an information system (e.g., computer technology) or network that resulted in an actual or potentially nationally relevant adverse effect on any of the three layers that constitute cyberspace, including information systems, networks, and/or the information residing therein (Harry et al., 2023, NIST).

As a first step, we use our datasets to characterize cyber incidents worldwide, showing that cybersecurity threats vary greatly across countries and their existence and proliferation depends on socio-economic and political factors. For example, we find that while high-income countries (HICs) are more likely to be targeted by financially motivated and exploitive incidents, non-HICs are mostly victims of politically motivated and disruptive incidents, mainly directed at the public administration. This evidence validates the claims that cybersecurity must be seen as more than just a technical aspect, but also as an economic matter (Clinton, 2023). These findings also confirm that governments should strive for customized strategies and measures based on the local context.

The first challenge in our study is the well-known problem of insufficient and fragmented data on cyber incidents (Anderson et al., 2013; Chen et al., 2023; Kigerl, 2016). To overcome this challenge, we work with two databases built by leveraging on open-source media articles published worldwide, to identify, collect, and classify disclosed cyber incidents. Media articles on cyber incidents exist in all countries (although in different volumes), provide open-source information, and minimize the risk of using commercial or reputational biased data (e.g., data from anti-virus companies, private firms, or governments). Therefore, using advanced AI and Machine Learning techniques, we scrape data from almost 5 million news articles published in 98 different languages from 2017 to 2022 and, after a process of translation, identification, and verification, we store approximately 27,000 cyber incidents spread in 179 countries in our Media Disclosed Cyber Events (MDCE) database. The advantage of this dataset is that it includes incidents from a large number of countries, including many low-income countries (LICs). However, this dataset does not classify incidents by industry, motive, or type. Therefore, we also use the Cyber Attacks Database collected by the Center of International and Security Studies (CISS) at the University of Maryland, which covers around 11,000 cyber incidents in around 150 countries from 2014 to 2023 and includes variables on incidents' characteristics (Harry and Gallagher, 2018). Although the CISS database collects incidents from major cybersecurity-related websites worldwide (e.g., cyber blogs and news sites recommended by practitioners, journalists, and academic researchers), it heavily relies on English-language sources and therefore, its coverage on LICs is limited. By working with both datasets, our aim is to test our hypotheses with two samples of what constitute the universe of cyber incidents. In general, both samples confirm that cyber incidents worldwide have been increasing in the last years, mostly since the COVID-19 pandemic and the war in Ukraine, and that, although the main targets are HICs, incidents directed towards non-HICs are increasing at fast rates.

To evaluate the link between cyber incidents and economic growth, we employ cross-country two-way fixed effects regression analysis. Our model accounts for variations in digital development levels and population size across countries. However, this methodology does not account for potential sources of endogeneity. Consequently, our findings should be interpreted as indicative correlations rather than a definitive proof of the impact of cyber incidents on economic performance. At a more granular level, we explore the causal effect of countries' cybersecurity commitments on the economic performance of industries. For this, we follow the cross-industry cross-country model developed by Rajan and Zingales (1996) and the model adaptations suggested by Ciccone and Papaioannou (2023). Cross-industry cross-country models have been widely used to study how the growth of industries in different countries depends on an interaction effect between country and industry characteristics. Since, more often than not, industry characteristics are unobserved in most countries, these models suggest approximating this variable with industry characteristics in a benchmark country. However, despite the popularity of these models in the economic literature, Ciccone and Papaioannou (2023) prove that this approach could yield biased estimates of the interaction (country and industry characteristic) effect depending on whether countries are similar across different characteristics. Following both papers, we apply a cross-industry cross-country model with an Instrument Variable adaptation to prove the positive effects of countries' higher cybersecurity commitments on industries' economic performance, given an approximated level of industries' exposure to cyber-attacks.

At the macroeconomic level, this study finds that an increase in cyber incidents occurring in low- and middle-income countries (non-High Income Countries or non-HICs) is, on average, statistically significantly correlated with lower GDP per capita, with a 95% confidence interval of 0.002% to 0.03%, while incidents in the education, finance, information, and public sectors are associated with up to 4 times bigger reductions in GDP. At a more granular level, this study finds that, given a level of cybersecurity vulnerability, industries grow faster in countries that have achieved higher levels of cybersecurity commitment, such as the creation of a national computer emergency response team (CIRT), a substantive cybercrime law, a national cybersecurity strategy, public cybersecurity awareness campaigns, the establishment of a cyber agency, among others.

This paper is organized as follows. Section 2 discusses the new field of Cybersecurity Economics by presenting the advances in the literature. The CISS and the MDCE datasets are described in Section 3, followed by the outline of our econometric methodology and reporting of the estimates in Section 4. A discussion of the results is undertaken in Section 5, including their wider policy ramifications and conclusions.

2. Literature Review

Created and manipulated by humans, cyberspace is an environment that allows for the interaction of over 5.3 billion individuals and 16 billion devices worldwide. Humans' centrality in the development of cyberspace, and our current levels of interconnectedness and dependency on the internet, make cybersecurity an economic matter with technical features more than just a technical issue. Thus, the relatively new field of Cybersecurity Economics seeks to re-think cybersecurity by understanding the economic decisions, factors, and processes involved in securing cyberspace. Studies in this field mainly focus on 1) the economic effects and the costs of cyber incidents, 2) the determinants of cybercrime, 3) how cybersecurity decisions are shaped by economic mechanisms (e.g., liability instruments), 4) market failures (e.g., spillover effects of cyber breaches) and the role of governments in the provision of a safe cyberspace, and 5) investment and insurance.

Scholarly works on the economic effects of cyber incidents mostly focus on U.S. firm-level effects (Amir et al., 2018; Garg, 2020; Kamiya et al., 2021, Lending et al.; Lin et al., 2020; Makridis; 2021; Piccotti and Wang 2022; Tosun 2021; Wang et al., 2022), analyzing the impact of a single large cyber incident (Crosignani et al., 2023), or the impact of cyber incidents on specific sectors like insurance (Woods et al. 2019; Woods et al. 2021), finance (Corbet and Gurdgiev, 2019; Jamilov et al., 2021, Kotidis and Schreft 2022), and transport (Austin and Withers, 2021). As a result, national and cross-industry evidence on the effects of cyber incidents is limited. However, pertinent to our findings, the mentioned studies suggest that cyber incidents can influence GDP via at least three channels. First, a cyber incident may diminish the productive capital and distort the production and delivery processes (Crosignani et al., 2023).[6,7] Second, cyber incidents could damage the reputation of firms, disincentivizing investments. For instance, in the short and medium-run, cyber breaches could impact the stock market by causing the equity price of publicly traded firms to decrease following a data breach announcement (Amir et al., 2018; Akey et al., 2021; Kamiya et al., 2021; Lending et al., 2018; Piccotti and Wang, 2022; Lin et al., 2020; Tosun, 2021). Third, a cyber incident can decrease the trust and the adoption of digital services affecting the digital economy

[6] For example, the Colonial Pipeline Ransomware attack in 2021 forced a U.S. energy company to shut down its entire fuel distribution pipeline for a few days, which threatened gasoline and jet fuel distribution across the U.S. east coast.
[7] A good example of that is the ransomware attack on MAERSK (2017), the Danish transport and logistics conglomerate, which halt the company shipment services globally for more than 10 days and forced the company to rebuild its entire IT infrastructure (a cost of around $300 million).

(Setiawan et al. 2018, Apau and Koranteng; 2019). Our paper builds on these findings and goes one step further by providing an estimate of the economic effects of cyber incidents and of cybersecurity commitments on the economic performance of nations and industries.

The expansion of a cyber-attack surface is rooted in the current coding processes that build vulnerabilities over vulnerabilities, as no one codes completely from scratch. However, technical aspects open vulnerabilities and the opportunities to attack, but it is socio-economic and political aspects that provide the motives and incentives for malicious actions. The studies on the determinants of cyber incidents confirm that although most cybersecurity efforts focus on technical aspects, cyber incidents are a socio-economic and political issue, offering an idea of why cyber-attacks exist and why they proliferate. For example, Chen et al., (2023) prove the socio-economic aspects of cybercrime by finding that poverty and income inequality incentivize cybercrime for illegal gains. Mezzour et al., (2014), Kigerls (2012), and Asal (2016) confirm this lucrative aspect of cybercrime by showing that economic growth indicates profitable victims. Kshetri (2019) add that unemployment also has a causal effect on cybercrime rates, especially in places with over-educated but under-employed computer experts (Hall and Ziemer, 2023; Onuora, 2017).

Another set of the literature studies how instruments like audits and certifications (e.g., product labels) increase product security and liability by informing consumers about their choices (Caven and Camp, 2023; Huang et al., 2024). These studies show that successful instruments depend on a variety of factors like understandability by the consumer, financial sponsorship by governments, tailored designs, etc.

The fourth set of studies on Cybersecurity Economics focus on market failures and government interventions, identifying four sources of failures in cybersecurity markets, misaligned vendors' incentives, asymmetric information, externalities, and network effects. Misaligned vendors' incentives are created by the unclarity on the returns of cybersecurity investments, and markets' tendency to reward first-movers (Huang et al., 2024). Information asymmetries arise because consumers of cybersecurity products have less information about the quality of the product than vendors, which creates downward pressure on both prices and quality (Anderson, 2013). Externalities are intrinsic to cyberspace; therefore, cyber incidents can spread beyond the victim and lead to even bigger security problems (Gandal, 2019). Finally, cybersecurity products exhibit direct and indirect network effects similar to IT, in which the value of a product increases with the number of consumers using it (direct effect), while also, increasing as complementary products and services become available (indirect effect) (Asghari et al., 2016). These four sources of market failures lead to government influences in cybersecurity markets through their roles as demanders of cybersecurity products, standards bodies, regulators and enforcers, and defenders (Rains, 2023).

The fifth and last literature group discusses issues related to cyber insurance like the quantification of cyber risk (Woods and Moore, 2019; Woods et al., 2021). Business leaders rank cybersecurity incidents as the top risk they face (World Economic Forum, 2022, 2023). However, quantifying cyber risk is an even greater challenge than counting cyber incidents, as measuring the risk entails estimating both successful and failed attacks and formalizing a probability function. This problem is exacerbated by the lack of consistent and reliable data on cyber-attacks.

3. Data

The main challenges for studying the economics of cybersecurity are the lack of reliable and unfragmented data and the absence of an agreed and shared mechanism on how to define, collect, and measure cyber incidents relevant at the national level (Aldasoro et al., 2020; Biener et al., 2015; Ho and Luong, 2022). For example, while some datasets include only "major incidents" (e.g., CSIS's[8] report on cyber incidents that incur in losses of over USD 1m), others report at a minimal level of granularity (e.g.,

[8] Center for Strategic and International Studies.

daily phishing attempts or vulnerabilities reports). Moreover, there is limited geographical coverage. For example, the most popular database used in cybersecurity economic studies, the Privacy Rights Clearinghouse dataset, only covers cyber incidents in the U.S. Finally, data from cybersecurity vendors or government agencies could face political or commercial biases (Amir et al., 2018; Howell and Burrus, 2020).

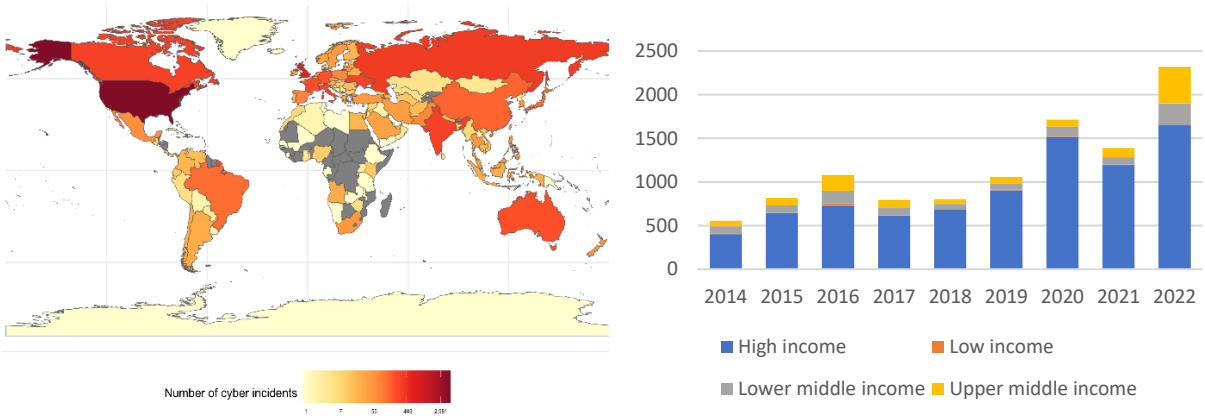
We deal with these challenges by focusing on non-personal cyber incidents reported by the news or relevant cybersecurity outlets, assuming that if an incident was covered by the media, more likely than not, it was an incident of considerable national relevancy. For this, we work with two databases on media disclosed cyber incidents, the CISS database by Harry and Gallagher (2018) and our Media-Disclosed Cyber Events database. We also work with data on other economic variables in the econometric modelling. The cyber incidents and economic variables databases are presented and discussed in this section.

3.1 Cyber Events Database (2014-2022)

The CISS's Cyber Events Database contains approximately 11,000 disclosed cyber incidents distributed in 156 countries from 2014 to 2022. It was built in three steps, 1) identification of relevant cybersecurity online resources (cyber blogs and news sites recommended by practitioners, journalists, and academic researchers; e.g., databreaches.net, redhotcyber.com), 2) execution of a script to make online requests and receive web data, and 3) manual validation of the events included in the database for quality assurance purposes and identification of incident's date, actor type, motive, threat actor country, targeted country, and industry.

The trend of cyber incidents shown by the CISS data shows that the highest peaks in cyber incidents in the last decade occurred during the COVID-19 pandemic (2019 and 2020), the Russia and Ukraine war (2022), and political activism in developing countries in 2016. Although the data show a significant concentration of incidents in HICs (mainly, in the U.S.), the highest growth rate is seen amongst UMICs and LMICs, mainly in LAC where incidents have increased at an annual growth rate of 30.2% versus the 19.5% global rate. Amongst non-HICs, omitting Russia and Ukraine, the countries with the highest number of cyber incidents are India, China, Brazil, Pakistan, and Mexico (Figure 2).[9]

Figure 2: Distribution of cyber incidents in the last decade



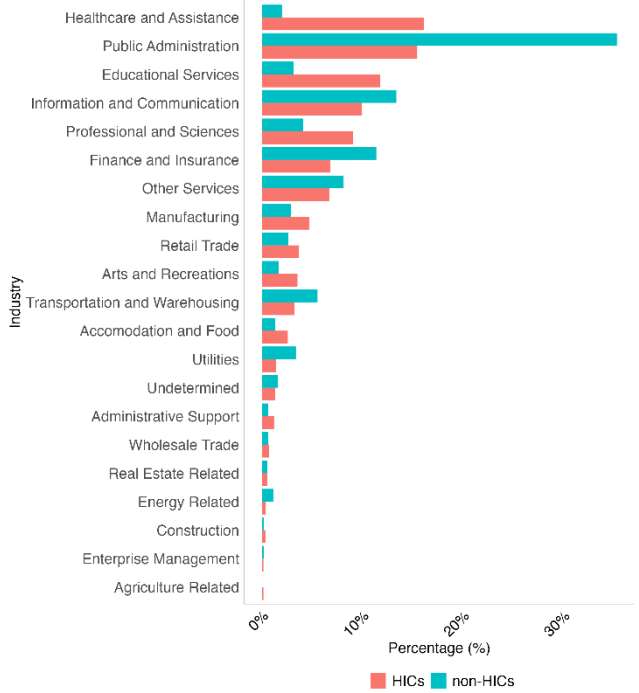
Source: Authors' own elaboration based on CISS Cyber Events Database

At the global level, the public administration ranks as the most attacked sector (19.4% of incidents in the last decade), followed by healthcare and social assistance (12.8%), information (11%), educational

[9] This database does not cover incidents for 66% of countries in SSA.

services (9.6%), and finance and insurance (7.9%).[10] However, as shown in Figure 3, attacked sectors vary by income group. Importantly, the public administration is the top target in non-HICs, with 35% of all non-HICs incidents targeting a public department or agency, followed by information (14%), finance and insurance (11%), other services (8%), transportation (6%), and utilities (3%). On the other hand, the distribution of incidents in HICs across sectors is less disperse, with healthcare at the top with 16.2% followed by the public administration and education with 15.4% and 11.7%, respectively.

Figure 3: Distribution of cyber incidents per sector by income group

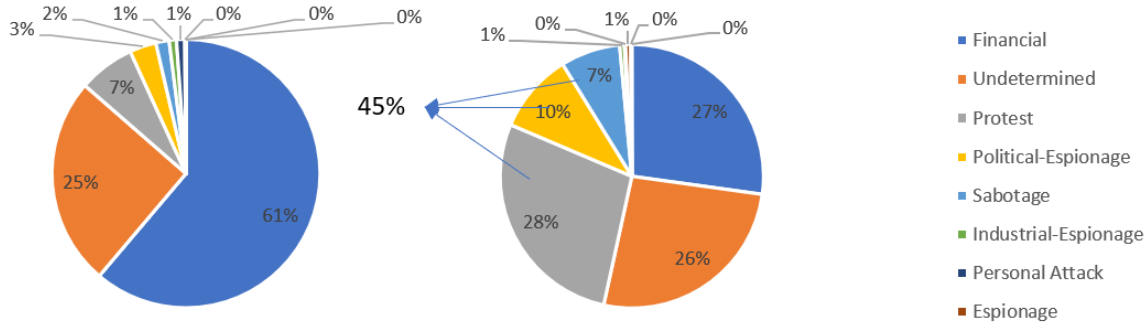


Source: Authors' own elaboration based on CISS Cyber Events Database 2014-2022.

Based on the categorization of incidents done by the CISS, HICs are mainly victims of financially motivated and exploitive cyber-attacks, whereas non-HICs are mostly targeted by politically driven and disruptive attacks like protests, espionage, and sabotage (Figure 4). Mainly, financial motives explain 54% of the cyber incidents worldwide, but 60% of the incidents in HICs, and 30% incidents in non-HICs. Protesting (e.g., hacktivists groups motivated by ideologies and the will to influence) is the second biggest motive worldwide explaining 11.2% of global incidents, but only 6.6% in HICs, and 28.1% in non-HICs. Across developing regions, while financial motivations are the top reason for carrying a cyber-attack in LAC (53%), EAP (48.4%), and SA (34.5%), political motivations explain most disclosed cyber incidents in SSA (41.1%) and MENA (28.2%)

[10] The CISS database classifies industries based on the North American Industry Classification System.

Figure 4: Distribution of motive of cyber incidents in HICs (left) and non-HICs (right)



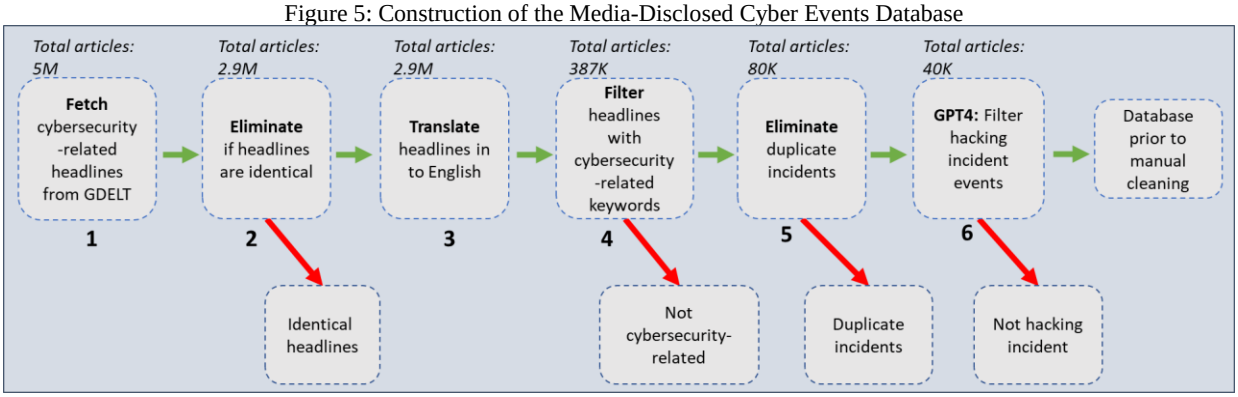
Source: Authors' own elaboration based on CISS Cyber Events Database 2014-2022.

The CISS defines two main types of cyber incidents, a "disruptive" incident impedes the target organization's normal operations, while an "exploitive" incident illicitly accesses or exfiltrates sensitive information such as personally identifiable information, classified information, or financial data.[11] There is symmetry between the motives and the type of cyber incidents as politically motivated adversaries will strive to disrupt systems and services rather than exploit existing data, whereas financial motivated incidents mainly strive to exploit data. Therefore, it is sensible to find more disruptive incidents in non-HICs (48%) and more exploitive incidents in HICs (53%)

3.2 Media-Disclosed Cyber Events Database (2017-2022)

Following a seven-steps data mining process, we construct an annual balanced panel of media-disclosed cyber incidents for 179 countries from 2017 to 2022. We expand on the fetching work by Kumar et al. (2016) who identified cyber incidents by using the GDELT[12] media reports that have an URL containing the word "cyber". We go a couple of steps further and identify cyber incidents by reading, translating, and cleaning the information from the headlines of over 5m media reports on cyber breaches reported around the world in 98 different languages and stored in the GDELT database (Figure 5). See more details on the construction of this database in Appendix 1 and 3.

Figure 5: Construction of the Media-Disclosed Cyber Events Database



Source: Authors' own elaboration.

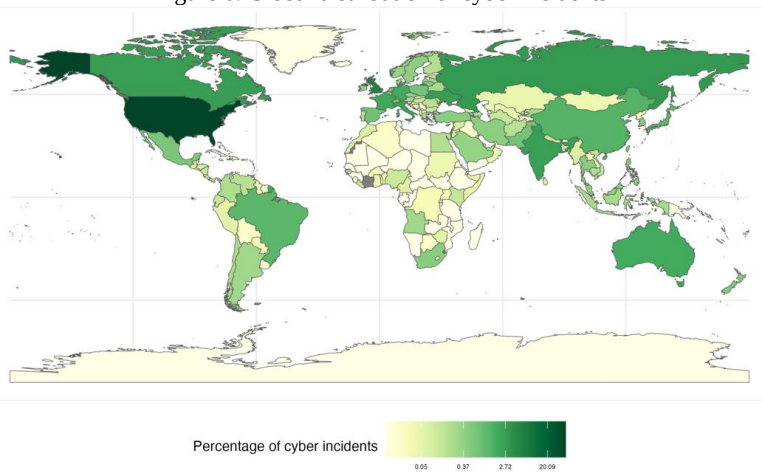
[11] The NIST provides a similar definition for these terms. Computer Network Exploitation (CNE) definition of NIST is "enabling operations and intelligence collection capabilities conducted through the use of computer networks to gather data from target or adversary information systems or networks". Disruption definition of NIST is "an unplanned event that causes the general system or major application to be inoperable for an unacceptable length of time (e.g., minor or extended power outage, extended unavailable network, or equipment or facility damage or destruction)."(see the NIST glossary, https://csrc.nist.gov/glossary/)

[12] Global Database of Events, Language, and Tone.

3.3 Comparison of the two cyber incidents databases

The MDCE database reports on 27,787 disclosed cyber incidents occurring in the lapse of 6 years in 179 countries while the CISS covers 11,100 incidents in 156 countries in 10 years. A sample examination shows that there is approximately a 50% overlap of incidents, which could indicate that while MDCE covers more incidents, some of them might be of a more micro nature (e.g., incidents directed at small firms), whereas the manual work in the production of the CISS database ensures that all the cyber incidents included are relevant at the national level. However, given that the CISS presents a geographical coverage gap, particularly in LICs, the MDCE database complements the global snapshot of cyber incidents (Figure 6). Moreover, the MDCE includes keywords on types of cybercrime such as "ransomware," "malware," "data breach," "DdoS," "exploit," "phishing," "spyware," and/or "hacking", which could allow, in future work, to study the effects of types of cybercrime.

Figure 6: Global distribution of cyber incidents



Source: Authors' own elaboration based on data imputation using the CISS and MDCE databases.

The increasing trend in cyber incidents worldwide is confirmed by both databases, although at varying rates. Mainly, since the COVID-19 pandemic, cyber incidents have been increasing at an annual growth rate between 4% (MDCE) and 24% (CISS). Consistent with our expectations, the share of HICs is lower in MDCE (69.5%) than in the CISS database (76.2%), which could signal on the importance of considering sources on multiple languages when building a media-based dataset.

3.4 Other data sources

The data on cyber incidents were merged to data on countries' economic performance and digitization, shown in Table 1.

Table 1: Descriptive statistics

Variable	Coverage	Source	N	#Countries	Mean	Std.
GDP per capita (constant 2015 USD)	1960-2022	World Governance Indicators	12780	258	10,923.5	17,581.8
% Internet users	2014-2021	ITU	1328	206	55.23	29.08
Population/1 mil	2014-2021	GSMA	1736	217	34.7	136

Mobile subscriptions/1 mil	2014-2021	GSMA	1556	205	38.98	142.96
GCI	2014	ITU	1,149	192	36.13	25.35
Industry level real gross value added/1 mil (USD)	2014-2021	UN	9,240	193	47,771.3	245,952.4

4 Methodology and estimation results

4.1 Baseline specification

Our baseline specification is

$$y_{it} = \alpha + \beta D_{it} + u_i + \theta_t + \Gamma Z_{it} + \epsilon_{it}, \quad (1)$$

where y_{it} is the natural logarithm of GDP per capita and D_{it} is the natural logarithm of the total number of disclosed cyber incidents of country i at time t. The country fixed effects term, u_i, absorbs time-invariant country characteristics like, for example, development level, culture, history, and response behaviors. The term θ_t represents the full set of time fixed effects that captures global developments that affect all countries similarly each year. The term Z_{it} is the vector of control variables that includes the percentage of internet users, the logarithm of total mobile subscriptions, and the logarithm of population size. Because the control variables may have their own effect on the treatment variable, we include three-period lagged values to minimize the direct relationship between the control variables and the treatment variable.[13] The error term ϵ_{it} includes all other time-varying unobservable shocks to GDP per capita. Finally, the coefficient of interest, β, represents the elasticity of GDP per capita with respect to cyber incidents.

Table 2 reports the two-way fixed effects (TWFE) coefficient estimates of the association between cyber incidents and GDP per capita separately for non-HICs, HICs, and all countries. In each panel, column (1) reports the TWFE regression, which may be biased if, for example, countries with greater digital development levels present higher economic growth rates despite receiving more incidents. Therefore, in column (2) we present the results controlling for population and digital development levels. The results show a statistically significant negative correlation between cyber incidents and GDP per capita in non-HICs and at the global level, the association is more significant after the inclusion of the control variables. Mainly, in non-HICS, a 1% increase in the number of cyber incidents is statistically significantly associated with a 0.0138% decrease in GDP per capita, while at the global level, a 1% increase in cyber incidents is associated with a 0.0125% decrease in GDP per capita.

Table 2: The effect of cyber incidents on GDP per capita in non-HICs, HICs, all countries OLS estimates

	A. Non-HICs		B. HICs		C. Global	
	(1)	(2)	(1)	(2)	(1)	(2)
Log(# incidents)	-0.00471	-0.0138**	-0.000292	-0.00872	-0.00454	-0.0125***
	(0.00674)	(0.00590)	(0.00520)	(0.00589)	(0.00439)	(0.00443)

[13] As documented by IBM (2022), it usually takes time until firms notice the presence of a cybersecurity incident. So, an incident that is disclosed in a year might have started in the previous years. If the changes in digital development level or population of a country drive a cybersecurity incident, controlling for these variables in the regression may create confounding bias. So, it is important to control for the lagged values of the control variables to minimize the contemporaneous relationship between the control variables and the treatment variable.

N	328	194	307	205	635	399
R^2	0.415	0.396	0.428	0.555	0.407	0.448
Controls	No	Yes	No	Yes	No	Yes

Each column reports a separate regression of the logarithm of GDP per capita on the logarithm of the total number of incidents in a country. Each regression controls for a full set of country and time-fixed effects. The control variables include three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions, and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in parentheses. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 level, respectively.

Heterogeneity analysis by industry following Equation 1 confirms that the sectors most economically impacted by cyber incidents in non-HICs are education, information, finance, and the public sector, with elasticities of 0.06, 0.029, 0.022, and 0.018, respectively (Table 3). In HICs, only cyber incidents directed to the education sector appeared as statistically significant.

Table 3: The effect of cyber incidents on GDP per capita in non-HICs by industry, OLS estimates

	Public (1)	Finance (2)	Education (3)	Information (4)
Log(# Incidents)	-0.0175**	-0.0219**	-0.0600***	-0.0292***
	(0.00803)	(0.0104)	(0.00889)	(0.00852)
N	125	69	25	71
R^2	0.504	0.539	0.979	0.535

Each column reports a separate regression of the logarithm of GDP per capita on the logarithm of total number of incidents specific to an industry in a country. Each regression controls for the full set of country and time fixed effects and control variables including three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in the parentheses. *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 level, respectively.

Table 4: The effect of cyber incidents on GDP per capita in HICs by industry, OLS estimates

	Public (1)	Finance (2)	Education (3)	Health (4)	Information (5)	Accommodation (6)
Log(#Incidents)	-0.00679	-0.0056	-0.0106**	0.00170	0.000696	0.0145
	(0.00535)	(0.00459)	(0.00497)	(0.00810)	(0.00537)	(0.0155)
N	126	90	59	63	95	42
R^2	0.532	0.642	0.741	0.504	0.672	0.802

Each column reports a separate regression of the logarithm of GDP per capita on the logarithm of total number of incidents specific to an industry in a country. Each regression controls for the full set of country and time fixed effects and control variables including three-period lagged values of the percentage of

internet users, the logarithm of total mobile subscriptions and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in the parentheses. *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 level, respectively.

In terms of the type of the cyber incident, all types—exploitive, disruptive, and mixed—are statistically significant associated with losses in GDP per capita amongst non-HICs. Mainly, a 1% increase in exploitive, disruptive, and mixed-type incidents are associated with reductions of 0.024%, 0.018%, and 0.038% in GDP per capita, respectively (Table 5).

Table 5: The effect of cyber incidents on GDP per capita in non-HICs by incident types, OLS estimates

	A. Non-HICs			B. HICs		
	Exploitive	Disruptive	Mixed	Exploitive	Disruptive	Mixed
	(1)	(2)	(3)	(1)	(2)	(3)
Log(#Incidents)	-0.0241***	-0.0179***	-0.0381**	-0.0108	0.00108	-0.000903
	(0.00827)	(0.00558)	(0.0156)	(0.00951)	(0.00480)	(0.00502)
N	148	117	65	161	153	93
R^2	0.371	0.528	0.609	0.571	0.568	0.799

Each column reports a separate regression of the logarithm of GDP per capita on the logarithm of the total number of incidents of a given type in a country. Each regression controls for a full set of country and time-fixed effects. The control variables include three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions, and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in parentheses. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 level, respectively.

4.2 Estimates with the Media-Disclosed Cyber Events Database

Applying the same model described in Equation 1 with the MDCE database confirms a negative correlation between GDP per capita and media-disclosed cyber events but not at a statistically significant level. These results could be explained by our initial impression that the MDCE database covers more micro-level incidents, or it just does not cover enough time periods.

Table 6: The effect of cyber incidents on GDP per capita, OLS estimates

	A. Non-HICs		B. HICs		C. Global	
	(1)	(2)	(1)	(2)	(1)	(2)
Log(# incidents)	-0.00764	-0.00923	-0.00627	-0.00279	-0.00705	-0.00796
	(0.0103)	(0.00704)	(0.00555)	(0.00447)	(0.00838)	(0.00560)
N	541	506	286	271	827	777
R^2	0.136	0.237	0.434	0.521	0.186	0.288
Controls	No	Yes	No	Yes	No	Yes

Each column reports a separate regression of the logarithm of GDP per capita on the logarithm of the total number of incidents in a country. Each regression controls for a full set of country and time-fixed effects. The control variables include three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions, and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in parentheses. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 level, respectively.

4.3 GMM estimates

We now estimate two different Arellano-Bond estimates of the effect of cyber incidents on GDP per capita in non-HICs using the CSIS data, one specification treats the count of cyber incidents as exogenous (GMM), and the other treats it as endogenous (GMM-IV). According to these results, one more cyber incident could be correlated with a decrease in GDP per capita in non-HICs between USD 2.4 and USD 2.7. GMM results for HICs and all countries can be found on Appendix 2.

Table 7: The effect of cyber incidents on GDP per capita in non-HICs, Arellano-Bond estimates

	GMM (1)	GMM-IV (2)
# Incidents	-2.708***	-2.403*
	(0.727)	(1.232)
GDP_{t-1}	0.436***	0.434***
	(0.144)	(0.148)
N	406	406
Countries	86	86
AR (2) test	0.728	0.721
Hansen overidentification test	0.307	0.419
Hansen exclusion test	0.719	0.419
Difference test	0.028	0.092
#Instruments	70	76

Each column reports a separate two-step Arellano-Bond regression of GDP per capita on the total number of incidents in a country. Column (1) constructs GMM instruments for the lagged value of GDP per capita while column (2) constructs instruments for both the lagged value of GDP per capita and the number of incidents. Each regression controls for a full set of country and time fixed effects, the lagged-value of GDP per capita, and control variables including three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in the parentheses. *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 level, respectively.

4.4 Cross-Country Cross-Industry Estimation

Following Rajan and Zingales (1996) we now explore the effect of countries' cybersecurity commitments on industries' growth rates, using the model

$$ValueAdded_{ij} = \alpha_i + \theta_j + \beta GCI_i * Vulnerability_j + share_{ij,2014} + v_{ij} \quad (2)$$

where $ValueAdded_{ij}$ is the compound growth rate[14] of industry *j* in country *i* between 2014 and 2021, α_i and θ_j represent country and industry fixed effects, respectively, GCI_i represents country i's level of cybersecurity commitment measured by the 2014 Global Cybersecurity Index, $Vulnerability_j$ is a measure of cybersecurity vulnerability (or, attractiveness to cybercriminals) of industry *j*, $share_{ij,2014}$ is the 2014 share of industry *j* in country *i*'s total value added, and v_{ij} is the industry and country specific error term. β, the main parameter of interest, captures the differential impact of a country's cybersecurity commitments

[14] The compound growth rate of an industry is computed using the formula $CAGR=(X_{2021}/X_{2014})^{1/7}-1$, where X_t is the real gross value added in year t.

on the growth rate of a sector given its level of cybersecurity vulnerability. Assuming that media coverage of cyber incidents, cybersecurity awareness, and reporting is more comprehensive in the U.S., we approximate an industry's vulnerability to cyber-attacks, $Vulnerability_j$, using the U.S.'s total number of disclosed cyber incidents per industry from 2017-2022. Because the U.S. is considered as the benchmark country, we exclude it from our estimation sample.

Ciccone and Papaioannou (2023), henceforth CP, show that the OLS estimates of cross-industry cross-country models could be biased when industry characteristics are proxied by a benchmark country. The bias could occur because the benchmark country could be a relatively poor proxy for some countries in the sample. In our case, if the GCI score is correlated with unobserved country-specific technological characteristics, this could lead to a biased estimate of β, if, for example, the U.S.-based vulnerability measure is a poor proxy for industries in countries with low GCI scores or lower technology. CP suggest that one can eliminate the bias by instrumenting the interaction term. The instrumental variable must be correlated with the interaction term but uncorrelated with the unobserved country and industry characteristics. We create such an instrument in four steps. First, we construct an industry-specific dummy variable, called "HighVulnerability", that indicates industries highly targeted by cyber incidents. In our sample, the agriculture and construction sectors are the least targeted industries, therefore, the dummy equals 0 in both cases (see Table 8). Second, we estimate the correlation between this "HighVulnerability" dummy and the growth rate of industries within a country to identify if highly targeted industries grow more or less in a country. Third, we create a dummy variable, D_i, that equals 1 if the estimated correlation is negative. That is, D_i equals 1 for 56 countries where highly targeted industries have lower growth rates. This new dummy variable varies only by country, while the error term captures characteristics that vary across industries and countries. For this reason, we expect D_i to be uncorrelated with the error term in equation (2). Finally, we interact D_i with $Vulnerability_j$. So, our instrument becomes $Z_{ij} = Di * Vulnerability_j$, which is both exogenous and correlated with the interaction term in Equation (2), meeting the requirements of a valid instrument.[15]

Table 8: Number of cyber-incidents per industry

	Disclosed cyber incidents
Agriculture, hunting, forestry, fishing (ISIC A-B)	4
Construction (ISIC F)	11
Manufacturing (ISIC D)	173
Mining, Quarrying, Utilities (ISIC C-E)	221
Transport, storage and communication (ISIC I)	79
Wholesale, retail trade, restaurants and hotels (ISIC G-H)	237

Table reports the total number of cyber incidents in the U.S. between 2014 and 2021 (based on the CISS database) for each industry in the United Nations' gross-value added database.

Table 9 provides first-time evidence of the importance of countries' cybersecurity commitments for industrial growth, by presenting the OLS and IV estimates of Equation 2. The results from both methods reveal a consistent and statistically significant positive link between the interaction term in Equation 2 and the industry-specific annual growth rates of the real gross value added. This suggests that higher levels of countries' cybersecurity commitments affect positively the economic performance of industries, given a level of cybersecurity threat. For the IV specification, the first-stage F-statistic exceeds the usual threshold, signifying the presence of a strong instrument. Moreover, the first-stage regression shows a negative and

[15] See Fiszbein et al. (2020) and Kukharsky (2020) for other recent cross-country cross-industry models that develop identification staretgies based on CP.

significant coefficient for the instrument, which implies that countries with lower GCI scores see greater growth in sectors that are less prone to cyber threats. This indicates a possible correlation between the unobserved component and the interaction term. Therefore, we confirm the OLS results with the IV estimate, which is also statistically significant and positive.

Table 9: The effect of cybersecurity commitments on the compound growth rate of real gross value added for industries

	(1) OLS	(2) IV
$GCI_i \times Vulnerability_j$	0.00018***	0.00092***
	(0.000054)	(0.00017)
First-stage		
F-statistic		23.48
Instrument		-12.29***
		(1.80)
N	848	848
# Clusters	183	183

The dependent variable is the compound growth rate of real gross value added in a sector between 2014 and 2021. Cyber vulnerability of an industry is proxied by the total number of incidents that occurred in the United States between 2017 and 2022. Each specification controls for a full set of country and industry fixed effects as well as the share of industry j in country i in 2014. In the IV specification, the instrument for the interaction term is a term that interacts cyber vulnerability of a sector with a dummy variable that is equal to 1 for countries where highly vulnerable industries have a lower growth rate than less vulnerable industries. Robust standard errors are provided in parentheses. The data source for the gross value added is the United Nations, while the source for cyber incidents data is the Cyber Events Database of the CISS. *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 level, respectively.

Next, we follow Rajan and Zingales (1996) and compare the actual (rather than estimated) effects of GCI on the growth rates of different industries. Table 10 contrasts the impact of cybersecurity commitments on the growth rates across various industrial sectors, stratified by countries with GCI scores below and above the median. Specifically, it shows the average residual growth rates for the real value added from 2014 to 2021, after accounting for industry and country-specific effects. For industries that experienced more cyber incidents, the residual growth rates in countries with lower GCI scores are in the negative range, with wholesale, retail trade, restaurants, and hospitality showing a decline of approximately -0.0406, and manufacturing exhibiting the most significant drop of around -0.8558. Conversely, in countries with higher GCI scores, these same sectors show positive growth, with manufacturing having the most substantial positive rate of approximately 0.7149. In the case of the least cyber-attacked industries, the pattern is reversed; countries with lower GCI scores show positive growth rates in construction and agriculture of 0.4524 and 1.4112, respectively, while these same industries in countries with higher GCI scores present negative growth rates.

Table 10: The effect of cyber-protection on the actual growth rate of real gross value added for industries, OLS estimates

	Countries with below-median GCI score	Countries with above-median GCI score
Most cyber-attacked industries		
Wholesale, retail trade, restaurants and hotels (ISIC G-H)	-0.04059	0.03146
Mining, Manufacturing, Utilities (ISIC C-E)	-0.30133	0.25937
Transport, storage and communication (ISIC I)	- 0.70363	0.58349
Manufacturing (ISIC D)	-0.85575	0.71493
Least cyber-attacked industries		
Construction (ISIC F)	0.45236	-0.39914
Agriculture, hunting, forestry, fishing (ISIC A-B)	1.41115	-1.35395

The table reports the average residual compound growth rate of real value added between 2014 and 2021 obtained after controlling for the industry and country fixed effects.

5. Conclusions

This paper presents first-time evidence on the role of cybersecurity on the economic performance of nations and industries. Using two novel datasets covering approximately 40,000 disclosed cyber incidents in about 180 countries from 2014 to 2022, and the results of a survey conducted to government officials of 194 countries on national cybersecurity commitments, we study the effect of cyber incidents on economic growth and the effect of cybersecurity commitments on industry performance. At the macroeconomic level, using a country and time fixed-effects model, we find evidence suggesting that cyber incidents are associated to losses in GDP per capita in developing nations. According to subsequent GMM results, we find that one cyber incident could be linked to GDP per capita losses in between USD 2.4 and USD 2.7 in non-HICs. Although we also find a negative correlation for developed nations, the results are not statistically significant. We argue that lower levels of cybersecurity commitments amongst developing nations could be playing a role in the effect that cyber incidents have on economic growth.

Then, we proceed to test a hypothesis about the positive effects of countries' cybersecurity commitments (e.g., legal, technical, and cooperation measures to control cybercrime) on industry economic performance. For this, we adapt the cross-industry cross-country model developed by Rajan and Zingales (1996), and following Ciccone and Papaioannou (2023) and others, we prove that for a given level of exposure to cyber incidents, an industry grows more in countries that have installed better cybersecurity commitments, controlling for country and time fixed effects like digitization and income levels. This constitute first-time evidence of the economic impact of cybersecurity capacity building.

Through an in-depth data analysis, we also document the characteristics of cyber incidents worldwide, paying particular attention at the differences between developed and developing countries. Consistent with the idea of characterizing cybersecurity as an economic, and not just a technical issue, we find that the characteristics of cyber incidents vary greatly across income groups and depend on socio-economic and political factors. For example, we find that while developed countries are mostly a target of financially motivated and disruptive incidents, developing nations are more likely to suffer politically motivated and exploitive cyber incidents. Moreover, while the public sector is the main victim in developing countries, the private sector, mainly, health, is the most targeted in developed countries. The implications of these findings could play a role on the design of tailored cybersecurity strategies and investment decisions.

Lastly, our literature review on the Economics of Cybersecurity highlights two important sources of knowledge gap on cybersecurity for non-HICs, 1) the centralization of cybersecurity research amongst developed countries (mainly, the U.S.), and 2) the lack of reliable and unfragmented data on cyber incidents, especially in non-HICs. To motivate the de-centralization of research in this area, governments and stakeholders could invest in data gathering initiatives that inform on the threat landscape of their respective countries. This could include collecting data on cyber incidents, vulnerabilities, costs of incidents, number of victims, amongst others. However, to do this, the international community must first agree on standard definitions and measures of key cybersecurity indicators such as nationally relevant cyber incidents. Especially since, in practical applications, parties are collecting data based on different definitions of cyber incidents. For example, while some parties identify a cyber incident based on the number of users compromised (e.g., the Chinese government considers a major incident one that compromises more than 10,000 users), others based this categorization on the monetary losses incurred by the incident (e.g., the U.S.'s CSIS considers a major cyber incident one that incurred losses of over USD 1m). A global agreement on definitions and metrics could help foster research and knowledge on this relatively new field.

Overall, our study derives value for cybersecurity investments from a risk-based approach that describes the threat landscape across income regions and the impact of cyber incidents on the economic performance of nations, thus, validating the belief that cybersecurity is an economic challenge of national importance, that requires more understanding from the social sciences viewpoint.

Bibliography

[1] Apau, R. and Koranteng, F.N., 2019. Impact of cybercrime and trust on the use of e-commerce technologies: An application of the theory of planned behavior. *International Journal of Cyber Criminology*, *13*(2).

[2] Akey, P., Lewellen, S., Liskovich, I., & Schiller, C. (2021) 'Hacking corporate reputations', *Rotman School of Management Working Paper*, (3143740).

[3] Aldasoro, I., Gambacorta, L., Giudici, P. and Leach, T. (2020) 'Operational and cyber risks in the financial sector'

[4] Amir, E., Levi, S., & Livne, T. (2018) 'Do firms underreport information on cyber-attacks? Evidence from capital markets', *Review of Accounting Studies*, 23, pp. 1177-1206.

[5] Anderson, R., Barton, C., Böhme, R., Clayton, R., Van Eeten, M.J., Levi, M., Moore, T. and Savage, S., 2013. Measuring the cost of cybercrime. The economics of information security and privacy, pp.265-300.

[6] Arellano, M. and Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, *58*(2), pp.277-297.

[7] Asal, V., Mauslein, J., Murdie, A., Young, J., Cousins, K. and Bronk, C., 2016. Repression, education, and politically motivated cyberattacks. *Journal of Global Security Studies*, *1*(3), pp.235-247.

[8] Asghari, H., van Eeten, M. and Bauer, J.M., 2016. Economics of cybersecurity. Handbook on the Economics of the Internet, pp.262-287.

[9] Austin, G. and Withers, G. (2021) 'Valuation of Reputation Damage for Transport Cyber Attack'.

[10] Biener, C., Eling, M. and Wirfs, J. H. (2015) 'Insurability of cyber risk: An empirical analysis', *The Geneva Papers on Risk and Insurance-Issues and Practice*, 40, pp. 131-158.

[11] Caven, P. and Camp, L.J., 2023. Towards a More Secure Ecosystem: Implications for Cybersecurity Labels and SBOMs. Available at SSRN 4527526.

[12] Chen, S., Hao, M., Ding, F., Jiang, D., Dong, J., Zhang, S., Guo, Q. and Gao, C., 2023. Exploring the global geography of cybercrime and its driving forces. *Humanities and Social Sciences Communications*, *10*(1), pp.1-10.

[13] Ciccone, A. and Papaioannou, E., 2010. Estimating cross-industry cross-country models using benchmark industry characteristics.

[14] Ciccone, A. and Papaioannou, E., 2023. Estimating cross-industry cross-country interaction models using benchmark industry characteristics. The Economic Journal, 133(649), pp.130-158.

[15] Clinton, L., 2023. Fixing American cybersecurity: Creating a strategic public-private partnership. Georgetown University Press.

[16] Corbet, S. and Gurdgiev, C. (2019) 'What the hack: Systematic risk contagion from cyber events', International Review of Financial Analysis, 65, 101386.

[17] Crosignani, M., Macchiavelli, M., & Silva, A. F. (2023) 'Pirates without borders: The propagation of cyberattacks through firms' supply chains', Journal of Financial Economics, 147(2), pp. 432-448.

[18] Crunchbase Report (2021). Report: The Rise of Global Cybersecurity Venture Funding, Crunchbase. https://about.crunchbase.com/cybersecurity-research-report-2021/. By

[19] eSentire & Cybersecurity Ventures (2022). Official Cybercrime Report. Available at: https://s3.ca-central-1.amazonaws.com/esentire-dot-com-assets/assets/resourcefiles/2022-Official-Cybercrime-Report.pdf (Accessed: 17 April 2023).
[20] Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., Baines, M., Celebi, O., Wenzek, G., Chaudhary, V. and Goyal, N., 2021. Beyond english-centric multilingual machine translation. *The Journal of Machine Learning Research*, *22*(1), pp.4839-4886.
[21] Fiszbein, M., Lafortune, J., Lewis, E.G. and Tessada, J., 2020. Powering Up Productivity: The Effects of Electrification on US Manufacturing. *NBER Working Paper*, (w28076).
[22] Gandal, N., 2019. Blockchain and the Law: The Rule of Code.
[23] Garg, P. (2020) 'Cybersecurity breaches and cash holdings: Spillover effect', Financial Management, 49(2), pp. 503-519
[24] Hall, T. and Ziemer, U., 2023. Cybercrime in Commonwealth West Africa and the Regional Cyber-Criminogenic Framework. *The Commonwealth Cybercrime Journal*, p.5.
[25] Harry, C., & Gallagher, N. (2018). Classifying cyber events. *Journal of Information Warfare*, 17(3), 17-31
[26] Hepfer, M., & Powell, T. C. (2020). Make cybersecurity a strategic asset. *MIT Sloan Management Review*, *62*(1), 40-45.
[27] Ho, H.T.N. and Luong, H.T., 2022. Research trends in cybercrime victimization during 2010–2020: a bibliometric analysis. SN Social Sciences, 2(1), p.4.
[28] Howell, C.J. and Burruss, G.W., 2020. Datasets for analysis of cybercrime. *The Palgrave handbook of international cybercrime and cyberdeviance*, pp.207-219.
[29] Huang, Z., Biczók, G. and Liu, M., 2024. Incentivizing Secure Software Development: The Role of Liability (Waiver) and Audit. arXiv preprint arXiv:2401.08476.
[30] IBM (2022). Cost of a Data Breach Report. Available at: https://www.ibm.com/reports/data-breach
[31] International Telecommunication Union (ITU) (2021). Global Cybersecurity Index (GCI). Available at: https://www.itu.int/dms_pub/itu-d/opb/str/D-STR-GCI.01-2021-PDF-E.pdf (Accessed: 19 January 2023).
[32] International Telecommunication Union, 2022. *Global Cybersecurity Index*. ITU.
[33] Jamilov, R., Rey, H. and Tahoun, A. (2021) 'The anatomy of cyber risk' (No. w28906), National Bureau of Economic Research.
[34] Kamiya, S., Kang, J. K., Kim, J., Milidonis, A., & Stulz, R. M. (2021) 'Risk management, firm reputation, and the impact of successful cyberattacks on target firms', *Journal of Financial Economics*, 139(3), pp. 719-749.
[35] Kigerl, A., 2016. Cyber Crime Nation Typologies: K-Means Clustering of Countries Based on Cyber Crime Rates. *International Journal of Cyber Criminology*, *10*(2).
[36] Kotidis, A., & Schreft, S. L. (2022). Cyberattacks and financial stability: Evidence from a natural experiment.
[37] Kshetri, N. (2019). Cybercrime and cybersecurity in Africa. *Journal of Global Information Technology Management*, *22*(2), 77-81.
[38] Kukharskyy, B., 2020. A tale of two property rights: Knowledge, physical assets, and multinational firm boundaries. *Journal of International Economics*, *122*, p.103262.

[39] Kumar, S., Benigni, M. and Carley, K.M., 2016, September. The impact of US cyber policies on cyber-attacks trend. In *2016 IEEE conference on intelligence and security informatics (ISI)* (pp. 181-186). IEEE.

[40] Lending, C., Minnick, K., & Schorno, P. J. (2018) 'Corporate governance, social responsibility, and data breaches', *Financial Review*, 53(2), pp. 413-455.

[41] Lin, Z., Sapp, T. R., Ulmer, J. R., & Parsa, R. (2020). Insider trading ahead of cyber breach announcements. *Journal of Financial Markets*, 50, 100527.

[42] Makridis, C. A. (2021) 'Do data breaches damage reputation? Evidence from 45 companies between 2002 and 2018', *Journal of Cybersecurity*, 7(1), tyab021.

[43] Mezzour, G., Carley, L. and Carley, K.M., 2014. Global mapping of cyber attacks. *Available at SSRN 2729302*.

[44] Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, pp.1417-1426.

[45] Onuora, A.C., Uche, D.C., Ogbunude, F.O. and Uwazuruike, F.O., 2017. The challenges of cybercrime in Nigeria: an overview. *AIPFU Journal of School of Sciences (AJSS)*, *1*(2), pp.6-11.

[46] Piccotti, L. R., & Wang, H. (2022) 'Informed trading in the options market surrounding data breaches', *Global Finance Journal*, 100774.

[47] Rains, T., 2023. Cybersecurity Threats, Malware Trends, and Strategies: Discover risk mitigation strategies for modern threats to your organization. Packt Publishing Ltd.

[48] Rajan, R.G. and Zingales, L., 1998. Financial dependence and growth. *The American Economic Review*, *88*(3), p.559.

[49] Reimers, N. and Gurevych, I., 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:19*

[50] Setiawan, N., Tarigan, V.E., Sari, P.B., Rossanty, Y., Nasution, M.D.T.P. and Siregar, I., 2018. Impact of cybercrime in e-business and trust. *Int. J. Civ. Eng. Technol*, *9*(7), pp.652-656.

[51] Tosun, O. K. (2021) 'Cyber-attacks and stock market activity', *International Review of Financial Analysis*, 76, 101795.

[52] Vergara, E. and Cakir, S., (2024)

[53] Wang, H. E., Wang, Q. E., & Wu, W. (2022) 'Short selling surrounding data breach announcements', *Finance Research Letters*, 47, 102690.

[54] World Economic Forum (2022). Cyber Resilience Index (CRI): Advancing Organizational Cyber Resilience. Available at: https://www.weforum.org/whitepapers/the-cyber-resilience-index-advancing-organizational-cyber-resilience/ (Accessed 21 January 2023).

[55] World Economic Forum (2023). Global Security Outlook 2023. Available at: https://www.weforum.org/reports/global-cybersecurity-outlook-2023 (Accessed 30 March 2023).

[56] Woods, D.W. and Moore, T. (2019). Does insurance have a future in governing cybersecurity?. IEEE Security & Privacy, 18(1), pp.21-27.

[57] Woods, D. W., Moore, T. and Simpson, A. C. (2021) 'The county fair cyber loss distribution: Drawing inferences from insurance prices', Digital Threats: Research and Practice, 2(2), pp. 1-21.

Appendix 1: Steps for construction of the Media Disclosed Cyber Incidents Database

Step 1: Through the execution of a script to receive web data from GDELT API, we identify and sort all articles that include in the headlines keywords liked "cyber", "ransomware," "malware," "data breach," "DdoS," "exploit," "phishing," "spyware," and/or "hacking." These terms encompass a wide variety of cybersecurity threats and provide a nuanced perspective on hacking incidents globally, ensuring comprehensive coverage of different types of hacking incidents. The first batch of extracted articles included 5 million articles. Then, we assign a temporary article's origin as the country of origin of the cyber event (see Step 6).

Step 2: We then drop identical duplicates which occurred if an event was recorded more than once in the database, possibly in different news outlets. Following the removal of identical duplicates, the database included 2.9 million articles.

Step 3: To address the issue of articles in a language other than English, we follow Fan et al. (2021) and use the m2m100_418M provided by Facebook to translate the articles from 98 languages. This model represents a state-of-the-art solution for multilingual machine translation tasks, thus providing high-quality translations of our dataset (see Appendix A.2 for a list of all translated languages present in our original database).

Step 4: Upon the successful translation of all headlines into English, the subsequent stage in our methodology focuses on further refining the data filtered by GDELT to ensure they are centered exclusively around cyber incidents. Thus, to enhance the precision of our data collection, we implement a secondary layer of filtration. This was done by employing lemmatization, a process that reduces words to their base or root form, allowing for the inclusion of various forms of a word. Consequently, the headlines of the articles were checked to ensure they contained combinations of the following attacks-related keywords: "ransomware", "malware", "data breach", "DdoS", "exploit", "phishing", "spyware", "hack" and "cyberattack." Employing lemmatization led to a database of 387,000 articles.

Step 5: Because the same incident can be published in multiple articles across various media platforms with a slightly different language, we seek to identify and group the articles that talk about a same incident (e.g., "DreamHost, web hosting company, blames powerful DdoS attack for online fraud", and "DreamHost, web host of controversial sites, hit by DdoS attack"). For this, we compare each pair of news headlines using a semantic similarity model and group the ones referring to the same cyber incident. In this process, we follow Reimers and Gurevych (2019) and use the Python package "sentence-transformers" and the "all-mpnet-base-v2" model built upon Microsoft's 'mpnet-base' model. This method was tested using a random testing sample of 3,487 pairs of headlines, which were manually categorized as duplicate articles. Applying the model to this sample, a threshold of 0.7 was established, above which pairs were considered to refer to the same event. The model demonstrated an accuracy score of 0.986 for this threshold in our test data. Upon this validation, the model was run across the entire dataset, comparing all combinations of article pairs, which led to approximately 80,000 unique articles.

Step 6: The final challenge in building our measure of cyber incidents is to distinguish between articles discussing a relevant cyber incident (e.g., 2017 Ukraine NotPetya ransomware attacks) and other news articles discussing general cybersecurity matters (e.g., best practices to follow when navigating the web, cyber hygiene tips, or celebrity related news).[15] For this, we use the GPT-4 API which, based on a specialized prompt defining a "Hacking Incident Detector", identified events as relevant incidents or not. The accuracy of the model was verified using a manually labeled test sample of 100 headlines. The model achieved an accuracy score of 0.876 in this test run. The directive for the model was as follows:
"*You are a Hacking Incident Detector GPT. Your purpose is to label news headlines as hacking incidents or not. The hacking incident definition is an occurrence that results in actual or potential jeopardy to the*

confidentiality, integrity, or availability of an information system or the information the system processes, stores, or transmits or that constitutes a violation or imminent threat of violation of security policies, security procedures, or acceptable use policies. The first requirement is that headlines are needed to mention occurred events, not opinions or conditionals. The second requirement is that please eliminate hacking incidents of individuals and celebrities. Hacking incidents must be mass events. Please take headlines as input and return True or False."

We also use GPT-4 to identify from the headlines another variable for the country of origin of the cyber incident. If the AI was able to identify a country of origin from reading the headlines, then we assumed that as the incident's country, if not, we used the variable generated in step 1. This step successfully finalized the data processing stage with approximately 40,000 valid articles reporting on cyber incidents.

Step 7: Finally, for accuracy, we manually check about 50% of the remaining articles to confirm that they discuss a real cyber incident, and that the country of origin is correctly identified. At the end of this data gathering process, we obtain 27,787 incidents. Approximately, 69.5% of the incidents are linked to HICs, followed by UMICs (16.9%), LMICs (13.1%), and LICs (0.2%), the rest (0.3%) were incidents that couldn't be linked to a country source.

Appendix 2: GMM results for HICs and all countries

Table 11: The effect of cyber incidents on GDP per capita in HIC, Arellano-Bond estimates.

	GMM (1)	GMM-IV (2)
# Cyber incidents	-2.653**	-1.878*
	(1.301)	(1.089)
GDP_{t-1}	1.063***	1.060***
	(0.0256)	(0.0282)
N	292	292
Countries	52	52
AR (2) test	0.340	0.340
Hansen overidentification test	0.442	0.702
Hansen exclusion test	0.565	0.872
Difference test	0.214	0.276
#Instruments	14	17

Each column reports a separate two-step Arellano-Bond regression of GDP per capita on the total number of incidents in a country. Column (1) constructs GMM instruments for the lagged value of GDP per capita while column (2) constructs instruments for both the lagged value of GDP per capita and the number of incidents. In this way, the table shows two different specifications, one treating the count of cyber incidents as exogenous (GMM), and the other treating them as endogenous (GMM-IV). Each regression controls for a full set of country and time fixed effects, the lagged-value of GDP per capita, and control variables including three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in the parentheses. *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 level, respectively.

Table 12: The effect of cyber incidents on GDP per capita global, Arellano-Bond estimates.

	GMM (1)	GMM-IV (2)
# Cyber incidents	-3.118***	-1.637**
	(1.142)	(0.644)
GDP_{t-1}	1.062***	1.063***
	(0.0215)	(0.0153)
N	786	786
Countries	139	139
AR (2) test	0.36	0.36
Hansen overidentification test	0.543	0.541
Hansen exclusion test	0.630	0.335
Difference test	0.269	0.797
#Instruments	14	17

Each column reports a separate two-step Arellano-Bond regression of GDP per capita on the total number of incidents in a country. Column (1) constructs GMM instruments for the lagged value of GDP per capita while column (2) constructs instruments for both the lagged value of GDP per capita and the number of incidents. In this way, the table shows two different specifications, one treating the count of cyber incidents as exogenous (GMM), and the other treating them as endogenous (GMM-IV). Each regression controls for a full set of country and time fixed effects, the lagged-value of GDP per capita, and control variables including three-period lagged values of the percentage of internet users, the logarithm of total mobile subscriptions and the logarithm of population. Robust standard errors, which are clustered at the country level, are given in the parentheses. *, **, and *** indicate significance at the 0.1, 0.05 and 0.01 level, respectively.

Appendix 3: Media-Disclosed Cyber Events Database

Table 13: List of languages translated to English when building the MDCE database

1. Afrikaans (af)	34. Gujarati (gu)	67. Persian (fa)
2. Albanian (sq)	35. Haitian; Haitian Creole (ht)	68. Polish (pl)
3. Amharic (am)	36. Hausa (ha)	69. Portuguese (pt)
4. Arabic (ar)	37. Hebrew (he)	70. Pushto; Pashto (ps)
5. Armenian (hy)	38. Hindi (hi)	71. Romanian; Moldavian (ro)
6. Asturian (ast)	39. Hungarian (hu)	72. Russian (ru)
7. Azerbaijani (az)	40. Icelandic (is)	73. Serbian (sr)
8. Bashkir (ba)	41. Igbo (ig)	74. Sindhi (sd)
9. Belarusian (be)	42. Iloko (ilo)	75. Sinhala; Sinhalese (si)
10. Bengali (bn)	43. Indonesian (id)	76. Slovak (sk)
11. Bosnian (bs)	44. Irish (ga)	77. Slovenian (sl)
12. Breton (br)	45. Italian (it)	78. Somali (so)
13. Bulgarian (bg)	46. Japanese (ja)	79. Spanish (es)
14. Burmese (my)	47. Javanese (jv)	80. Swahili (sw)
15. Catalan; Valencian (ca)	48. Kannada (kn)	81. Swati (ss)
16. Cebuano (ceb)	49. Kazakh (kk)	82. Swedish (sv)
17. Central Khmer (km)	50. Korean (ko)	83. Tagalog (tl)
18. Chinese (zh)	51. Lao (lo)	84. Tamil (ta)
19. Croatian (hr)	52. Latvian (lv)	85. Thai (th)
20. Czech (cs)	53. Lithuanian (lt)	86. Tswana (tn)
21. Danish (da)	54. Luxembourgish; Letzeburgesch (lb)	87. Turkish (tr)
22. Dutch; Flemish (nl)	55. Macedonian (mk)	88. Ukrainian (uk)
23. English (en)	56. Malagasy (mg)	89. Urdu (ur)
24. Estonian (et)	57. Malay (ms)	90. Uzbek (uz)
25. Finnish (fi)	58. Malayalam (ml)	91. Vietnamese (vi)
26. French (fr)	59. Marathi (mr)	92. Welsh (cy)
27. Fulah (ff)	60. Mongolian (mn)	93. Western Frisian (fy)
28. Galician (gl)	61. Nepali (ne)	94. Wolof (wo)
29. Ganda (lg)	62. Northern Sotho (ns)	95. Xhosa (xh)
30. Gaelic; Scottish Gaelic (gd)	63. Norwegian (no)	96. Yiddish (yi)
31. Georgian (ka)	64. Occitan (post 1500) (oc)	97. Yoruba (yo)
32. German (de)	65. Oriya (or)	98. Zulu (zu)
33. Greek (el)	66. Panjabi; Punjabi (pa)	

We follow Fan et al. (2020) and use the m2m100_418M provided by Facebook to translate media reports provided in this table to English.