Appendix A: Codebook

Category	Variable	Description and Codes
publications	pub_n	provide a sequential item number that will serve as publication ID
	first_author	provide the last name and first name first letter of the first author
	publication_title	provide full publication title
	publication_url	publication url
	publication_year	year when the paper was published
	publication_venue	venue name where the paper was published (e.g., name of the journal or conference proceedings)
	publication_venue_type	venue type where the paper was published (e.g., conference/ symposium proceedings (P), journal(J), repository(R), Other (O))
	abstract	copy publication abstract
	empirical	does the paper have an empirical component that requires the use of data (y/n)?
	research_focus	describe paper focus, using abstract and keywords as the primary source
	analysis_type_1	describe the primary kind of analysis that is done in this paper using the following categories:
		- prototype and evaluation: creating / building a system that involves hardware and / or software components, usually its something that has an architecture
		- algorithm development and testing: proposing a new algorithm or an addition to an existing algorithm, usually its a computer program written in a programming language or pseudocode
		- machine learning application: using machine learning to learn something about a domain or application area (can be applied to networks, human behavior, threats, etc.)
		- vulnerability analysis: examine software, standards, or other components for vulnerability
		- network scan: scanning internal or external networks to identify certain configurations, vulnerabilities, or patterns
		- survey: collecting data by asking people questions or examining documents to extract thematic information
		- conceptual model: introducing a description or representation of components using logic, mathematical equations, or other abstract language
		- statistical analysis: reporting descriptive statistics, testing hypotheses using statistical methods (regression, etc.)
	analysis_type_2	describe the secondary kind of analysis (using categories from analysis_type_1)
	analysis_type_3	describe a third type of analysis if it has a significant role in the publication
	analysis_tools	list computational tools that were used in the analysis
	comment	additional comments and details
datasets	pub_n	provide item number of the publication where the dataset came from
	dataset_name	provide dataset name if available

	origin	describe data origin
		- existing: uses an existing data from a third party
		- simulated: data collected from an experimental setup or simulated environment
		- synthetic: data generated to be similar to existing datasets
		- collected: data collected from natural, real-world environments
		- other: data that does not fit four above categories
	dataset_description	brief description of the data
	dataset_availability	describe how the dataset is made available
		- publicly available: shared publicly by authors without restrictions
		- restricted access: available with restrictions (registration, payment, etc.)
		- available upon request
		- partially available: part of the dataset is available publicly
		- not available: dataset is not available and not access conditions are provided
	link	link to where the dataset is available or shared
	type of link to data	describe the type of reference to the data (no url or citation; url to dataset; url to repository or website; doi, citation, no doi; broken link)
	analytics_availability	describe whether the researchers share their code (yes, no)
	analytics_link	link to where the code is available or shared
	zheng_taxonomy	select a taxonomy category: attack-related; defender artifacts; end users and orgs; internet characteristics
	sauerwein_taxonomy	select a taxonomy category: vulnerability; threat; countermeasure; attack; risk; asset
researchers	last_name	researcher last name
	first_name	researcher first name
	title	researcher title
	email	researcher email
	webpage	link to CV or webpage with information if available
	department	researcher department
	home_institution	researcher home institution
	research focus	describe researcher focus in cybersecurity research
	pub_n	list all publications from "publications" category that this researcher authored
	comment	record any observations or notes that are related to this name
first_authors	pub_n	provide item number of the publication where the dataset came from
	total_n_authors	provide total number of authors for this publication
	first_author	first author name
	first_name	first author last name
	title_at_pub	position at the time of publication
	current_title	current position
	gender	gender
	email	first author email

	web_link	link to CV or webpage with information if available
	department	first author department
	pub_institution	first author pub_institution
	home_institution	first author home_institution
	research focus	first author research focus