SOMETIMES THREE RIGHTS REALLY DO MAKE A WRONG: MEASURING CYBERSECURITY AND SIMPSON'S PARADOX

Eric Jardine
Assistant Professor of Political Science, Virginia Tech, and CIGI Fellow
ejardine@vt.edu

Overview

- Worsening aggregate cybersecurity trends
- The argument
- Assumptions/conditions
- The data
- The findings
- What it all means

Worsening Aggregate Trends

Figure 1. Breaches and Stolen Records Over Time

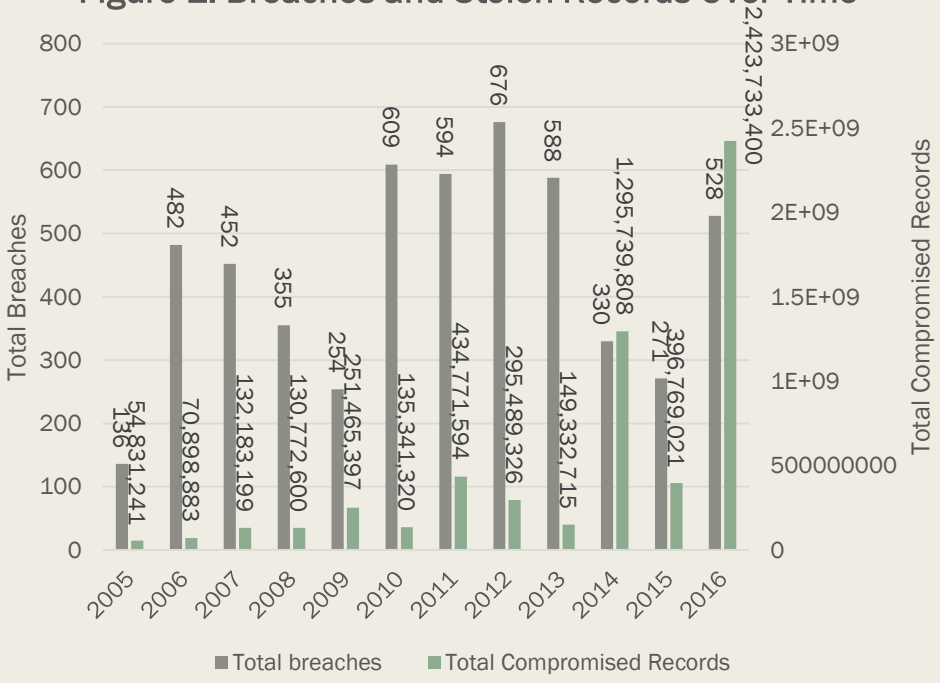
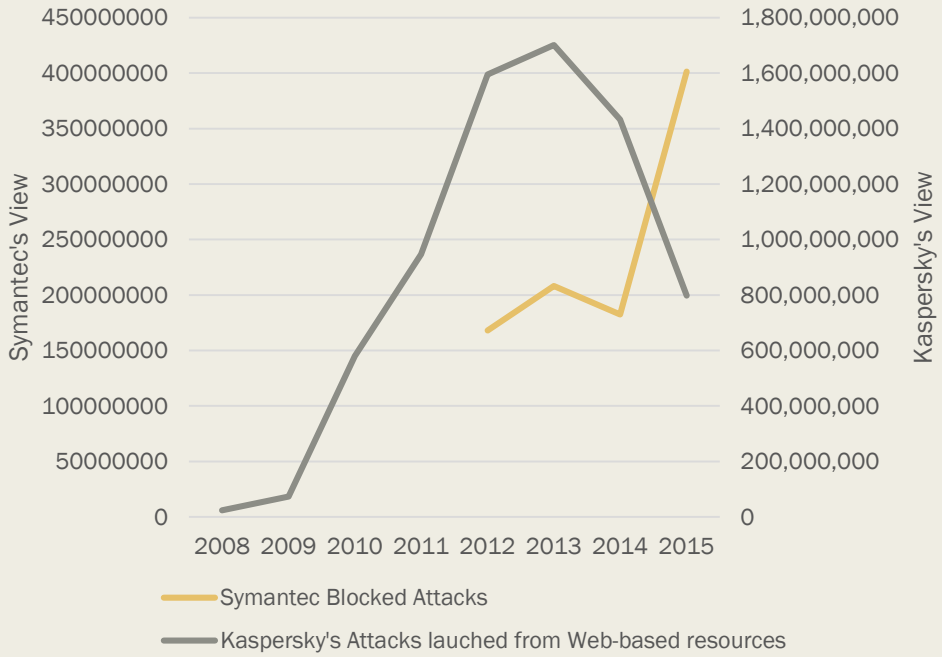


Figure 2. The View From IT Security Firms



The Argument

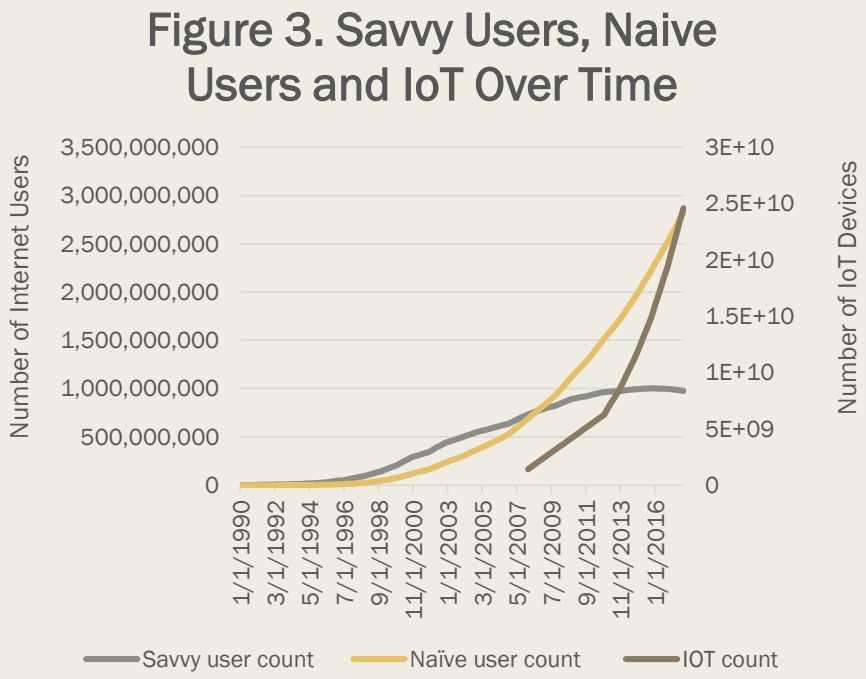
- First, an example: aggregation bias in *“A Nation at Risk”*—a report of the National Commission on Excellence in Education.
- What we see in aggregate statistics could, paradoxically, be based upon opposite trends.
- In other words, beware of aggregation bias in cybersecurity metrics, as they could be misleading us to believe that radical and disjunctive policy is needed when in fact the situation is getting better.
- If three conditions hold, I show that Simpson’s Paradox or aggregation bias could easily emerge to cloud our view.

Assumptions/Conditions

- The population of potentially hackable points online can be divided into groups.
- For example, 1) savvy users; 2) naive users; and 3) IoT devices.
- The likelihood that the groups will be hacked can be rank ordered
- For example, the likelihood of being hacked is: savvy users < naïve users < IoT
- The most hackable groups grow the fastest.
- For example, the various groups grow: savvy users < naïve users < IoT

The Data

- The Real…



- The Simulated…

Table 3. Cybersecurity in a Gaussian World

	Improvement Rate	Min	Max	Mean	Median	Standard Deviation
Savvy Users	0.1	0.29	0.50	0.40	0.40	0.044
Naïve Users	0.1	0.44	0.62	0.54	0.53	0.042
IoT	0.1	0.59	0.73	0.66	0.66	0.032

Table 5. Cybersecurity in a Power Law World

	Xmin	α	Min	Max	Mean	Median	Standard Deviation
Savvy Users	Savvy Users * 0.4	3	0.33	1.02	0.50	0.45	0.15
Naïve Users	Naïve Users * 0.5	2.5	0.39	5.25	0.77	0.57	0.57
IoT Device	Devices * 0.6	2	0.45	6.60	0.98	0.67	1.03

The Findings

- If cybersecurity is normally distributed…

Table 4. Simpson's Paradox Over 1,000 Iterations of the Data (a Gaussian World)

	Hack Proportion (1990-1996)	Hack Proportion (1997-2003)	Hack Proportion (2004-2010)	Hack Proportion (2011-2017)	Over Sample Percentage Point Change
Savvy Users	44%	41%	38%	35%	-9%
Naïve Users	59%	56%	53%	50%	-9%
IOT	N/A	N/A	67%	65%	-2%
Aggregate Trend	45%	45%	52%	61%	16%

- If cybersecurity is fat tailed…

Table 6. Simpson's Paradox Over 1,000 Iterations of the Data (Power Law World)

	Hack Proportion (1990-1996)	Hack Proportion (1997-2003)	Hack Proportion (2004-2010)	Hack Proportion (2011-2017)	Over Sample Percentage Point Change
Savvy Users	57.8%	53.8%	50.9%	49.0%	-8.8%
Naïve Users	80%	73.3%	69.1%	65.2%	-14.8%
IoT	N/A	N/A	89.8%	84.7%	-5.1%
Aggregate Trend	59.9%	59.3%	67.4%	81.2%	21.3%

What it all means

- Three lessons:
  1. *Take all aggregate cybersecurity statistics with a grain of salt;*
  2. *When devising metrics and collecting data, look for 'lurking confounders';*
  3. *Radically disjunctive policy might not be warranted.*

- And a call…
  1. *We need more, and more finely grained, data.*