



PRIVACY CON

FEDERAL TRADE COMMISSION

📍 DC // 1.14.16



Session 4: Economics of Privacy & Security



PRIVACY CON
FEDERAL TRADE COMMISSION

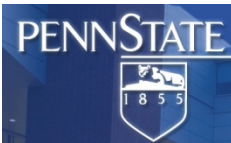
Jens Grossklags

Pennsylvania State University

An Empirical Study of Web Vulnerability Discovery Ecosystems

Co-authors: Mingyi Zhao, Peng Liu (Pennsylvania State University)





An Empirical Study of Web Vulnerability Discovery Ecosystems

Mingyi Zhao, Jens Grossklags, Peng Liu

Pennsylvania State University



Bug Bounty History



(1995)



Bug Bounty Platforms (Web Vulnerability Discovery Ecosystems)

HackerOne



Wooyun



Motivation and Approach

- **Motivation:** Detailed studies of Web vulnerability discovery ecosystems are absent
 - Debate on the impact of bug bounties for web security
 - Policy: e.g., limits on legality of vulnerability research
- **Approach:** Empirically study characteristics, trajectories and impact of two representative ecosystems
 - Stakeholders: Companies/organizations, white hats, black hats, public, policymakers, bounty platform providers etc.
 - Focus of this presentation: **Perspective of companies and organizations**



Web Vulnerability Data

	HackerOne	Wooyun
HQ	US & NL	China
Founded	2013-11	2010-07
# Vulnerabilities	10,997	64,134
# White hats	1,653	7,744
Participation Model	Organization-initiated	White hat-initiated
Bounty Level	Avg. \$424 (Various)	Very Low
Disclosure	Partial	Full
Data Collected	Bounty Amount Response Timeline	Vulnerability Type Severity

(Data until 2015-07)



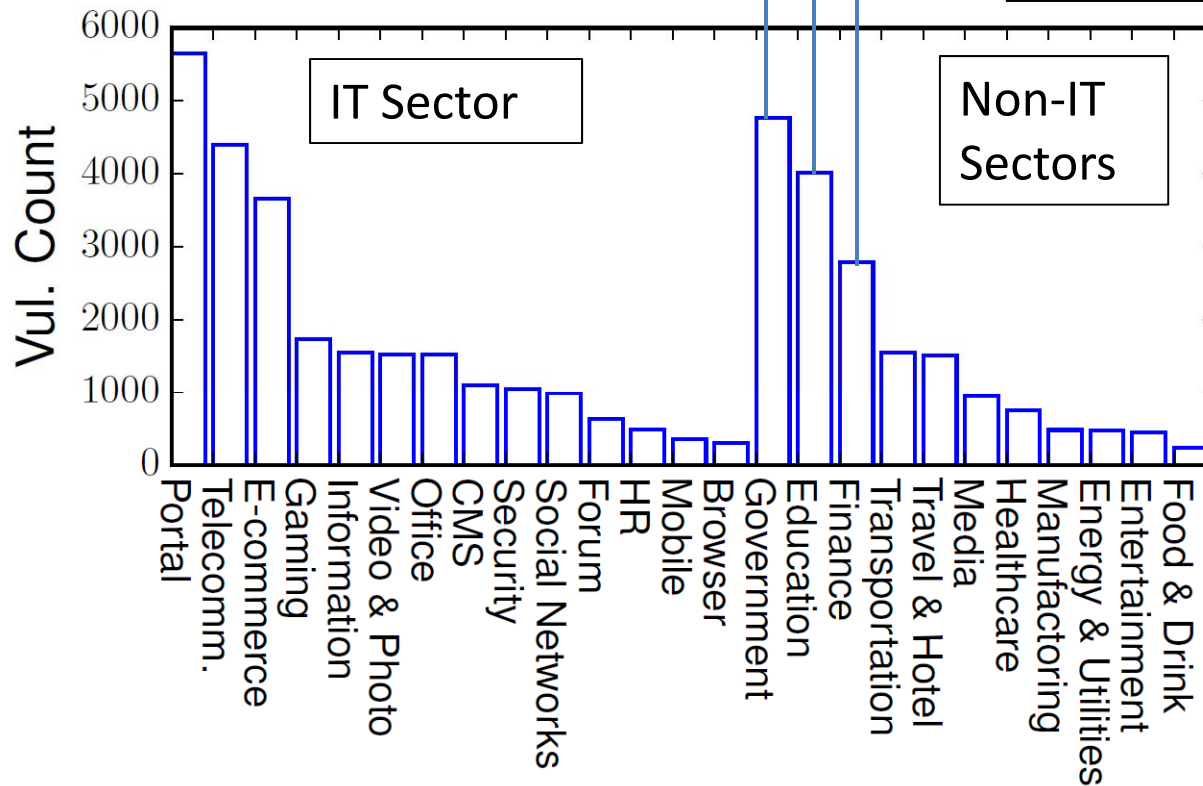
Participation by Organizations



Organization Types

- **HackerOne:** 99 organizations
 - All IT companies
 - Social networking, security, bitcoin ...
- **Wooyun:** 17328 organizations

	# Orgs	# Vuln
Gov	3179	4772
Edu	1457	4017
Fin	1040	2794



Takeaways: Participation

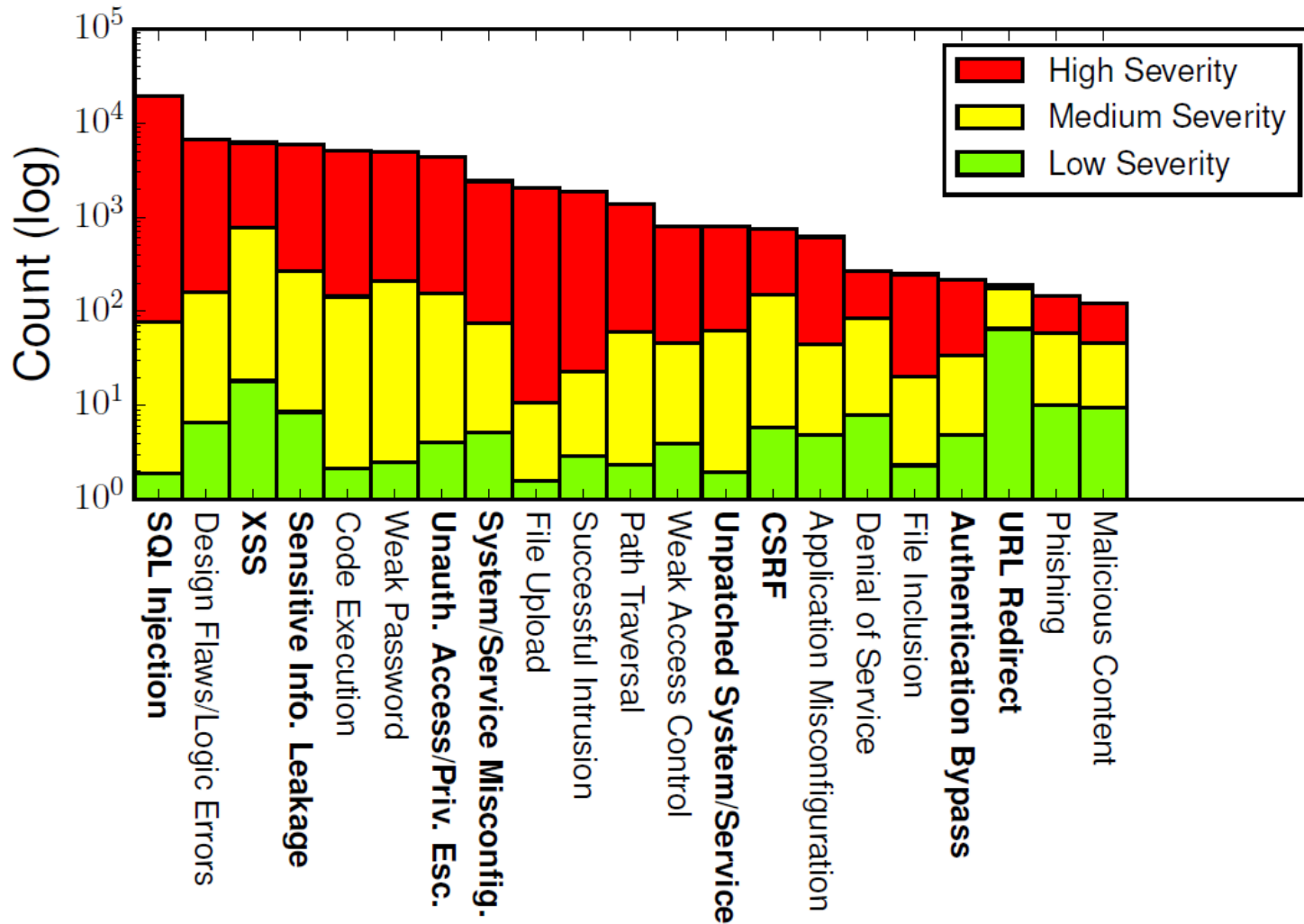
- The white hat-initiated model (Wooyun) achieves a much broader coverage of organizations
 - Less constraints for targeting organizations with web security issues
 - Growing size and diversity of the white hat community
- More limited participation under the organization-initiated model (HackerOne)
 - Raises question about ways to increase participation by companies and white hats



Types of Vulnerabilities & Severity



Types & Severity - Wooyun

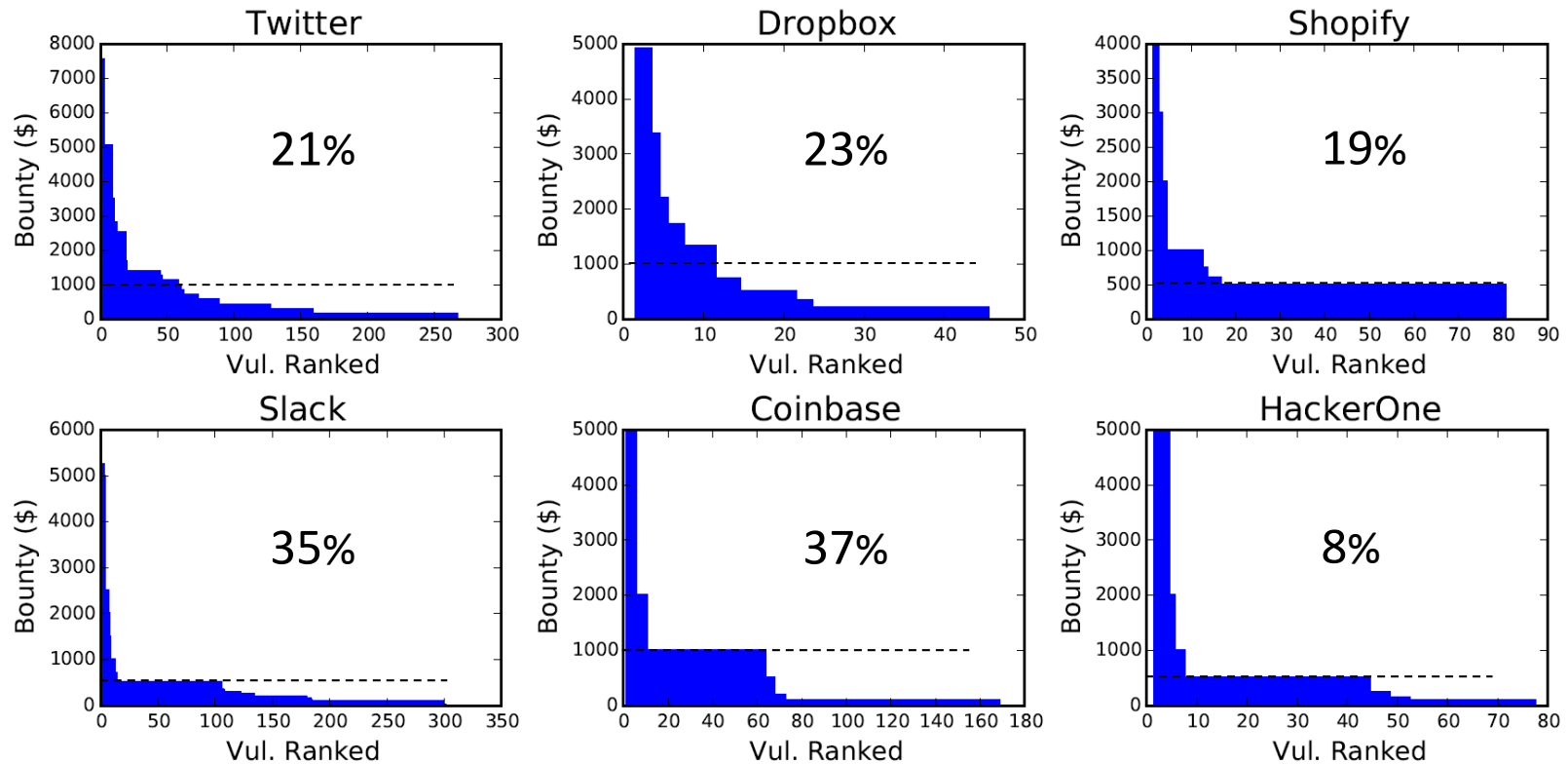


High: 44%
 Med: 40%
 Low: 16%

* OWASP TOP 10 in bold font

Severity – HackerOne

- Infer medium and high severity percentage from bounty distribution and policy statement



Takeaways: Types & Severity

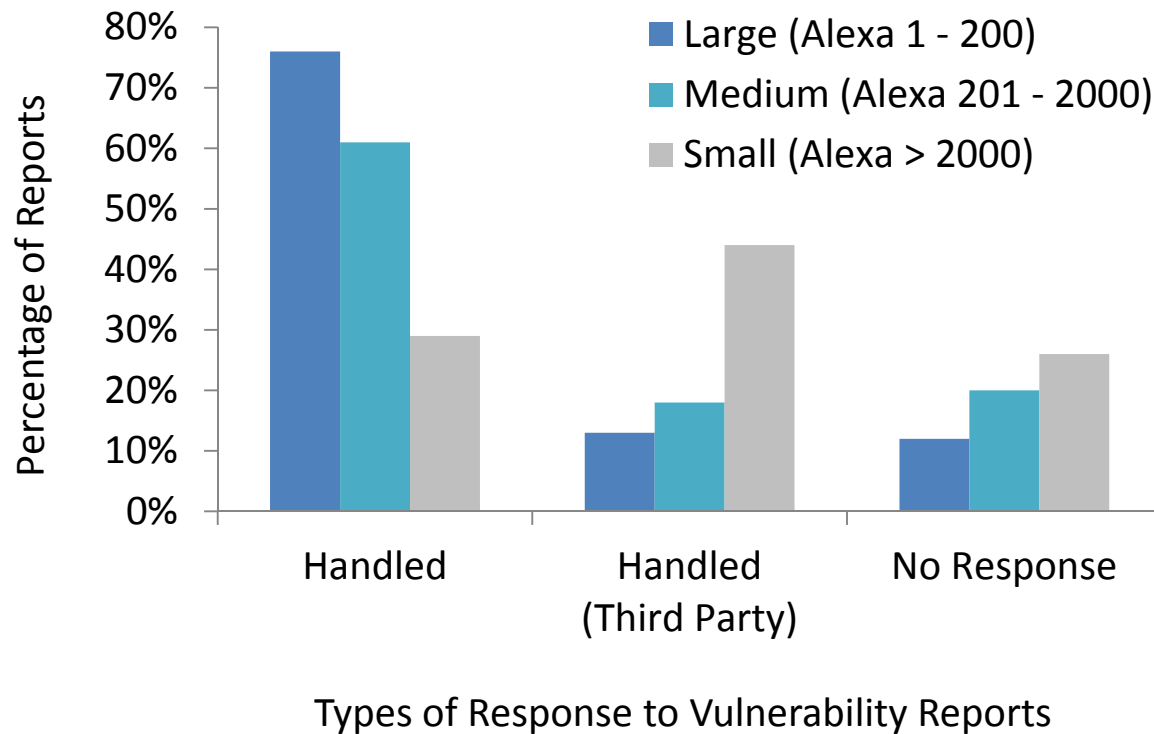
- White hats make considerable contributions
 - Broad range of vulnerability types
 - Significant percentage of medium/high severity reports
- White hat-initiated model (Wooyun) harvests potential of the community more comprehensively
 - Occasional contributors perform as a group almost as well as top white hats in terms of finding high severity issues
 - Can organizations properly handle all reports?



Response Behaviors by Organizations

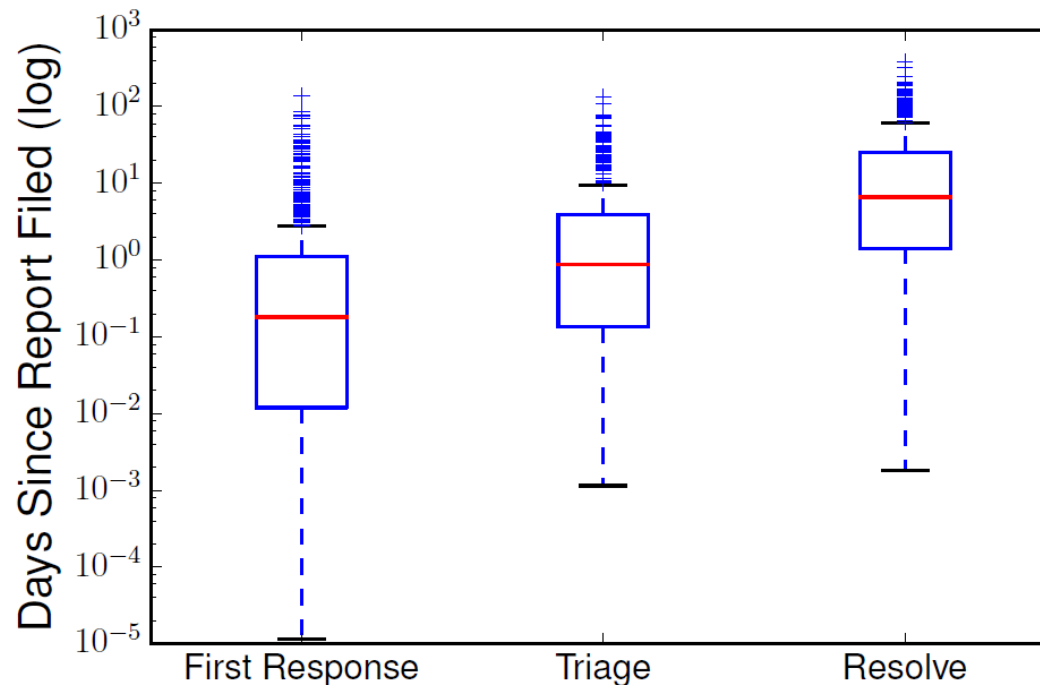
Response - Wooyun

- Segmenting organizations by Alexa rank (i.e., popularity) reveals differences in response patterns



Response - HackerOne

- Organization-initiated programs handle most reports and have quick response times
 - First response time median: 4.5 hours
 - 90% of the disclosed reports were resolved in 30 days





Takeaways: Response

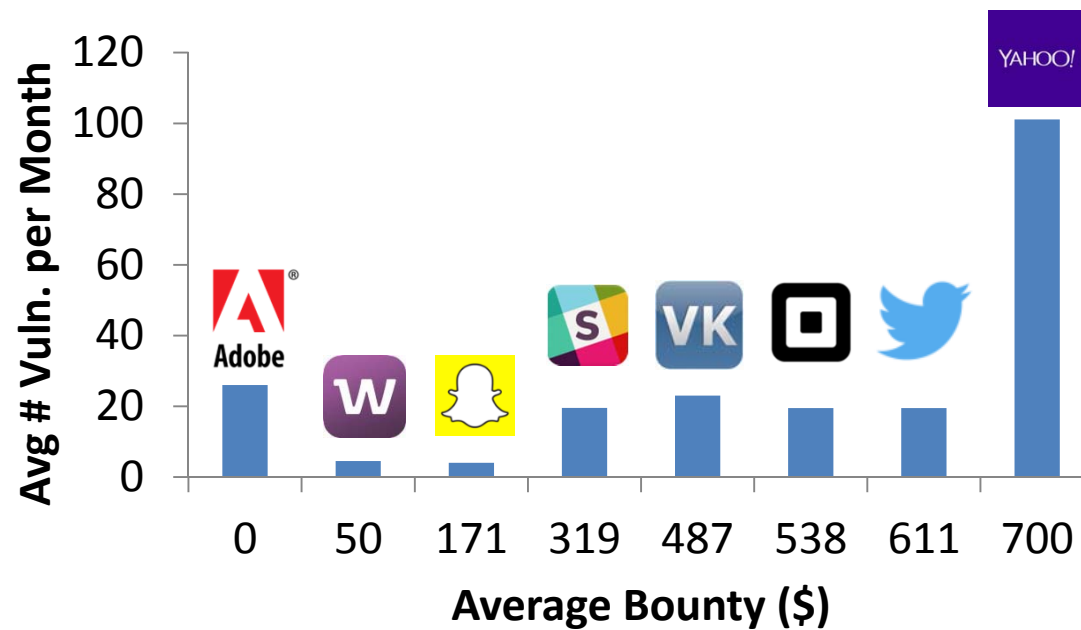
- Wooyun: Many organizations are not prepared!
 - Particularly smaller, less popular websites (Alexa > 2000)
- White hat-initiated model may increase risk for unprepared organizations
 - Vulnerabilities are published after 45 days; vulnerabilities with no response (unhandled) could still be exploitable
 - Balance trade-off between applying pressure and reasonable expectations for response by small companies



Cost and Impact of Bounties

Bounty Structure - HackerOne

- Different levels of bounties:



Impact of bounties?

Impact of Bounties

- Regression methodology
- Dependent variable:
 - Average # reports per month
- Independent variables:
 - Average bounty
 - Alexa rank
 - Platform manpower (time-weighted # white hats / # org.)



Regression Analysis Results

VARIABLES	(1) # Vuln.	(2) # Vuln.	(3) # Vuln.
Avg. Bounty (R_i)	0.04*** (0.01)	0.03*** (0.01)	0.03*** (0.01)
Alexa [log] (A_i)		-2.52* (1.20)	-2.70** (1.21)
Platform Manpower (M_i)			10.54 (10.14)
Constant	3.21* (1.88)	16.12** (6.39)	-133.05 (143.66)
R-squared	0.35	0.39	0.40

Standard errors in parentheses
 *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

+\$100 ~
 +3 vuln./month
 More popular
 ↓
 More attention
 More complex
 ↓
 More reports

Takeaways: Bounties

- While HackerOne puts focus on monetary compensation of white hats, we still observe many contributions (20% of all reports) to programs without bounties (33% of all programs)
 - Pay-nothing is a viable approach
- However, higher bounty amount is associated with considerable increase of number of vulnerability reports

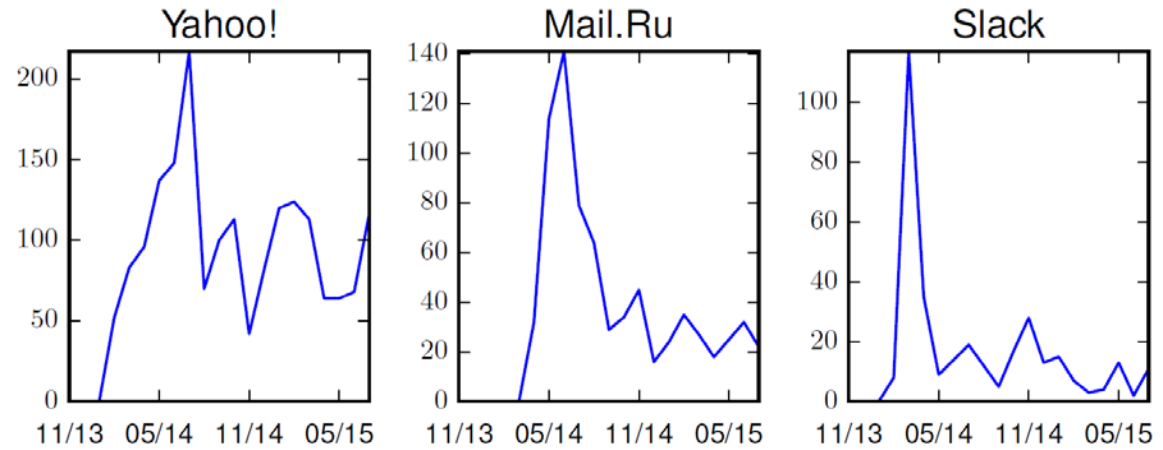


Security Improvements

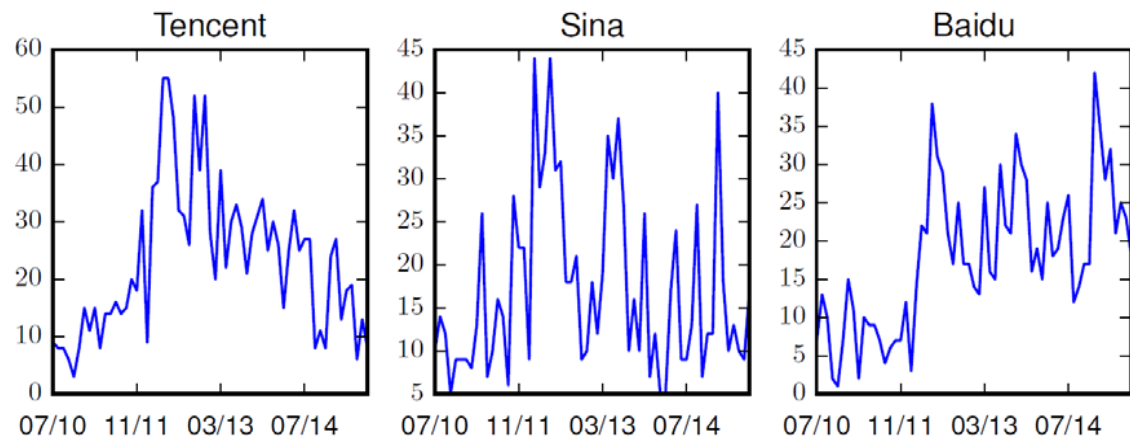


Vulnerability Trend: Data

- HackerOne:



- Wooyun:





Statistical Trend Test

- Laplace Test
 - Used in previous vulnerability study (Ozment, 2006)
 - Criteria: ≥ 4 months and ≥ 50 reports

	Decrease	Increase	No Trend
HackerOne	32	8	9
Wooyun	11	81	17

Takeaways: Trends

- Despite/because monetary incentives on HackerOne:
Fewer vulnerabilities are found over time
 - Indicative of improved web security of participating IT companies
 - Initial spike: With sufficient incentive, many vulnerabilities which likely were known/existed before launch are reported when the program opens
- Opposing trend for Wooyun programs
 - Likely worse integration between bug bounty program and SDL

Thank you.

- Comparison between different Web vulnerability ecosystems provides unique opportunities to study effectiveness of policies and practices
 - Many more results in the paper(s)
- Jury is still out about which participation model offers the most important benefits



Veronica Marotta

Carnegie Mellon University

Alessandro Acquisti

Carnegie Mellon University

Who Benefits from Targeted Advertising?

Co-author: Kaifu Zhang (Carnegie Mellon University)



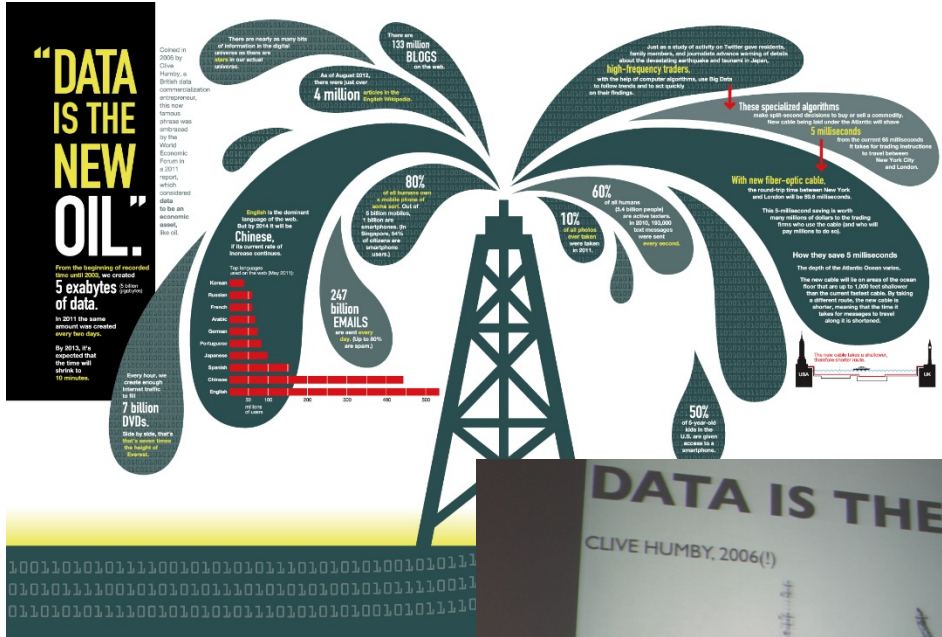
Who Benefits From Targeted Advertising?

Veronica Marotta, Kaifu Zhang, Alessandro Acquisti

Carnegie Mellon University

Federal Trade Commission

PrivacyCon 2016



Data is the new oil.

We see in data the same transformative, wealth-creating power that 19th-century visionaries once sensed in the crude black ooze trapped underground.

If "crude" data can be extracted, refined, and piped to where it can impact decisions in real time, its value will soar. And if data can be properly shared across an entire ecosystem and made accessible in the places where analytics are most useful, then it will become a true game changer, altering the way we live, work, learn, and play.

Source: Cisco IBSG, 2012. #DataInMotion

Data is the new

DATA IS THE NEW OIL

CLIVE HUMBY, 2006(!)

DATA is the new oil

Data is the new Oil

Image Sourced from coverrealestate.com



Personal information is the lifeblood of the Internet

Loss of privacy is the price to pay for the benefits of big data

Sharing personal data is an economic win-win

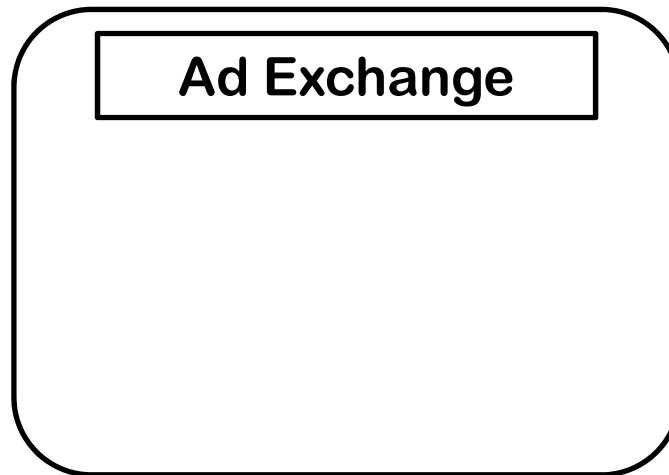
Who Benefits from Targeted Advertising?

- To what extent availability of more and more precise information about consumers leads to:
 - An increase in total welfare?
 - A change of allocation of benefits between different stakeholders?
 - Firms
 - Consumers
 - Intermediaries (i.e. ad exchanges)

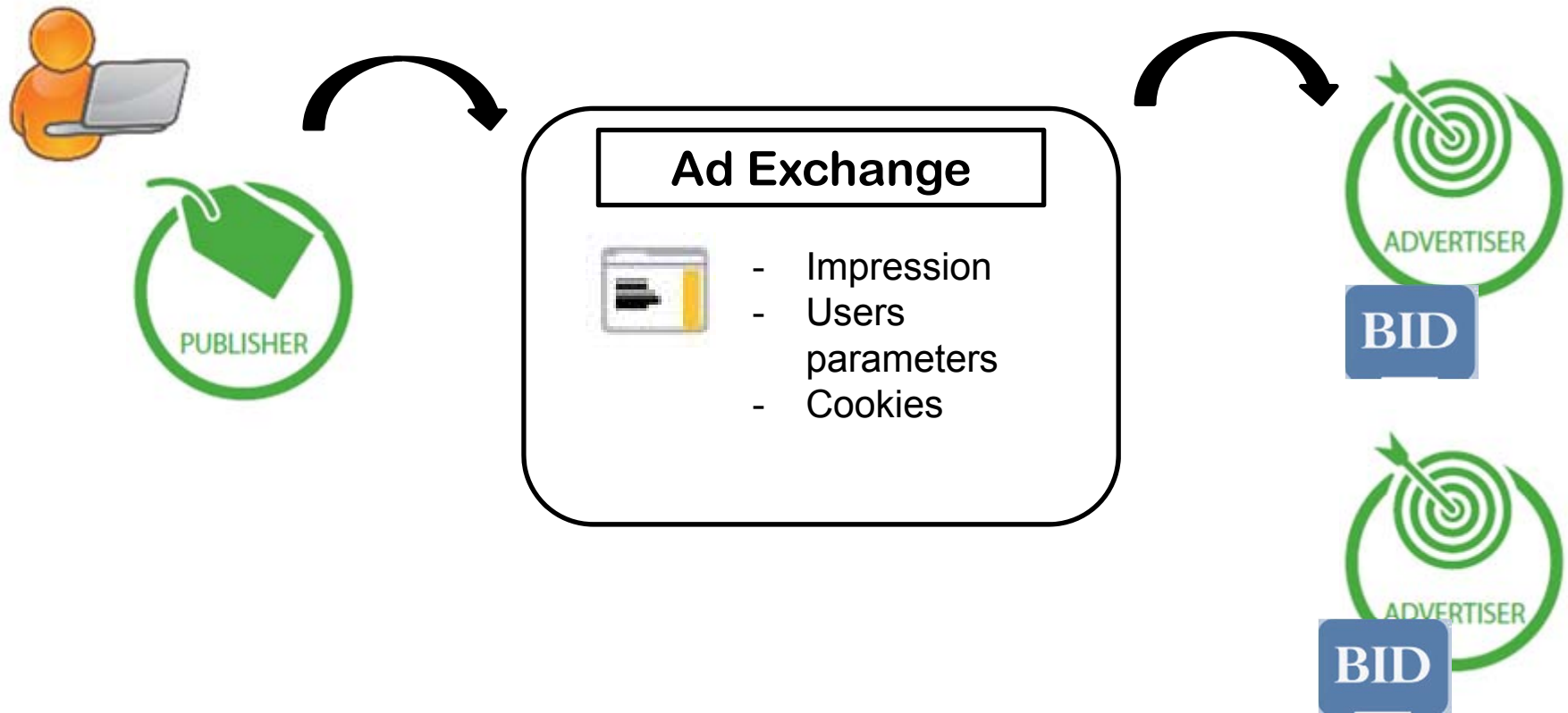
Methodology

- Multi-stage, 3-players model of online targeted advertising
- Compare scenarios that differ in the type and amount of consumer's information available during the targeting process
- Account for the role of the intermediary (the ad exchange) in the advertising ecosystem
- Focus on "Real-Time Bidding" (RBT)

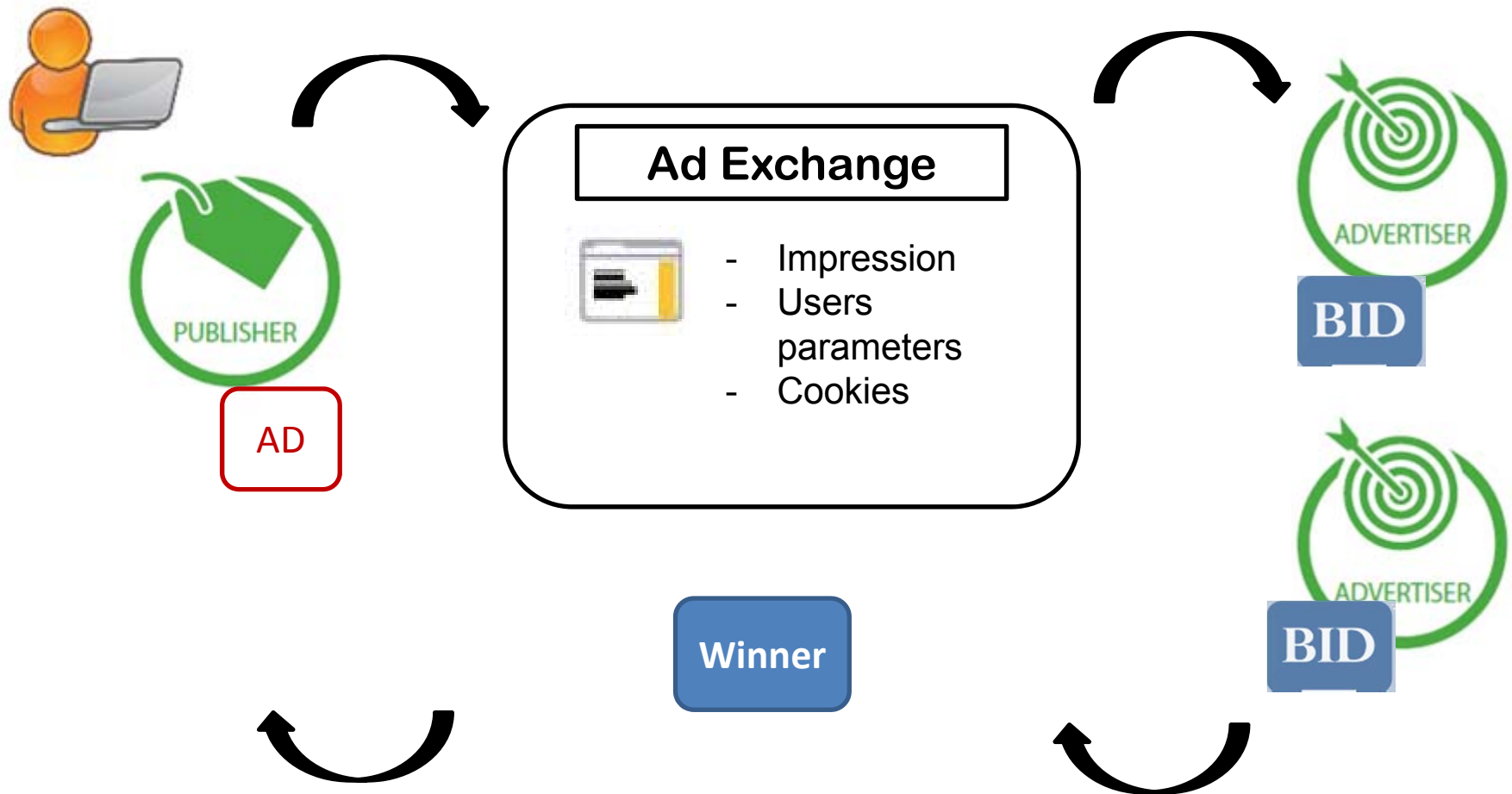
Advertisers are Bidding for Consumers



Advertisers are Bidding for Consumers



Advertisers are Bidding for Consumers



The Model: Basic Setting

1. Firms (the Advertisers):

- Profit-Maximizer
- Cannot target consumers directly

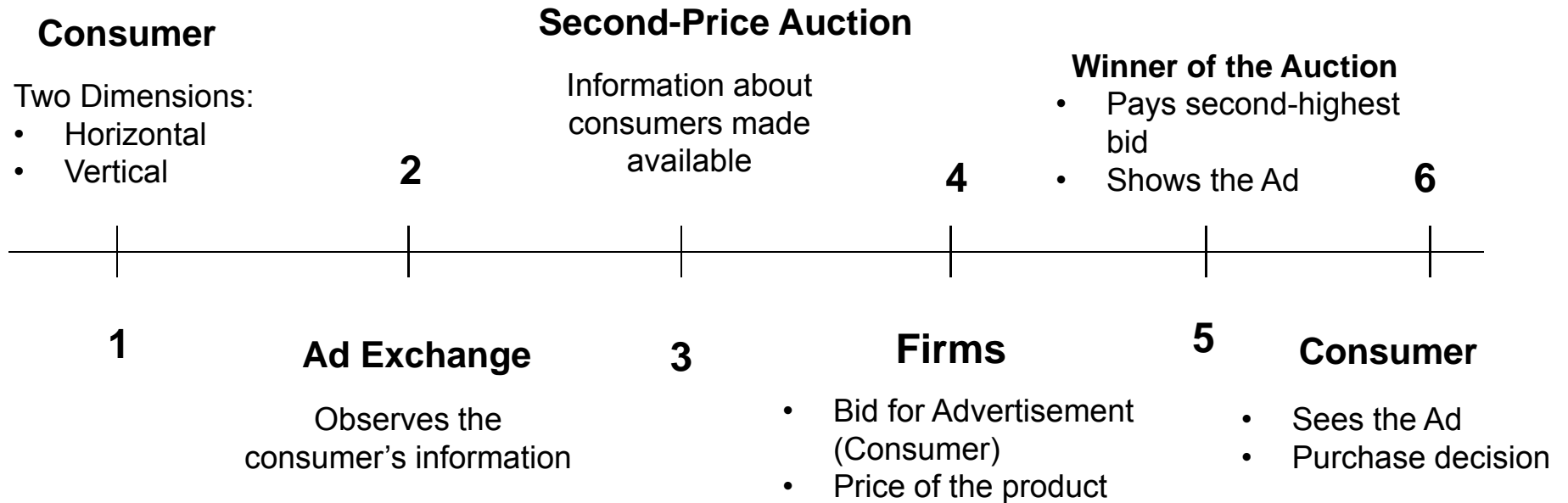
2. Intermediary (the Ad Exchange):

- Profit-Maximizer
- Runs auctions for advertisements' allocation

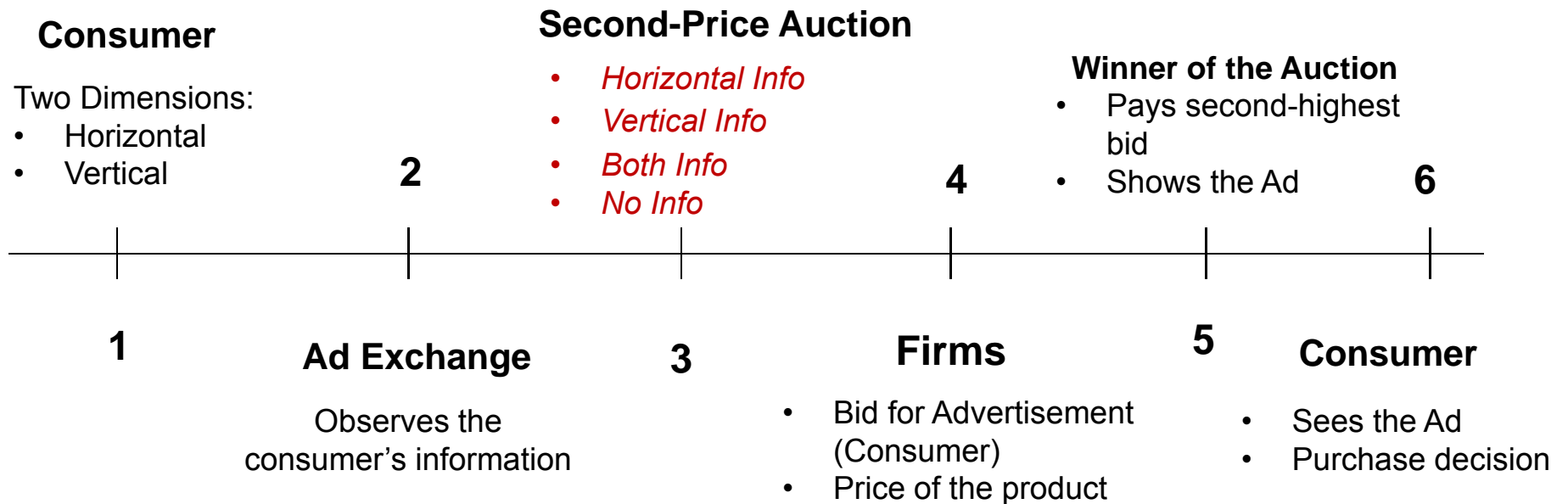
3. Consumers:

- Have product preference, but need to find seller
- Differ along two dimensions: horizontal (brand preference) and vertical (purchase power)

The Model: Sequence of Events



The Model: Sequence of Events



Analysis

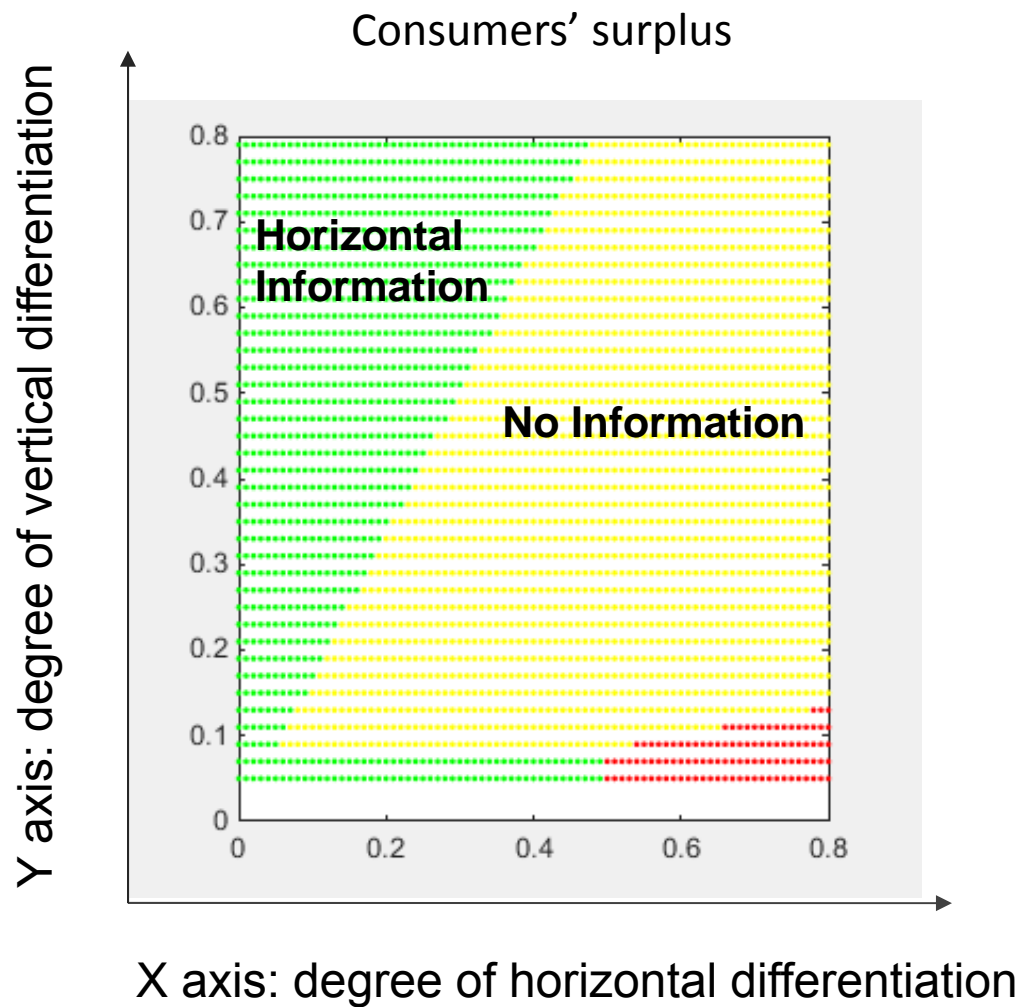
1. For each scenario, we derive:

- Firm's bidding strategy (for advertisement)
- Firm's pricing strategy (for product being advertised)
- Intermediary's profit
- Consumer's choice

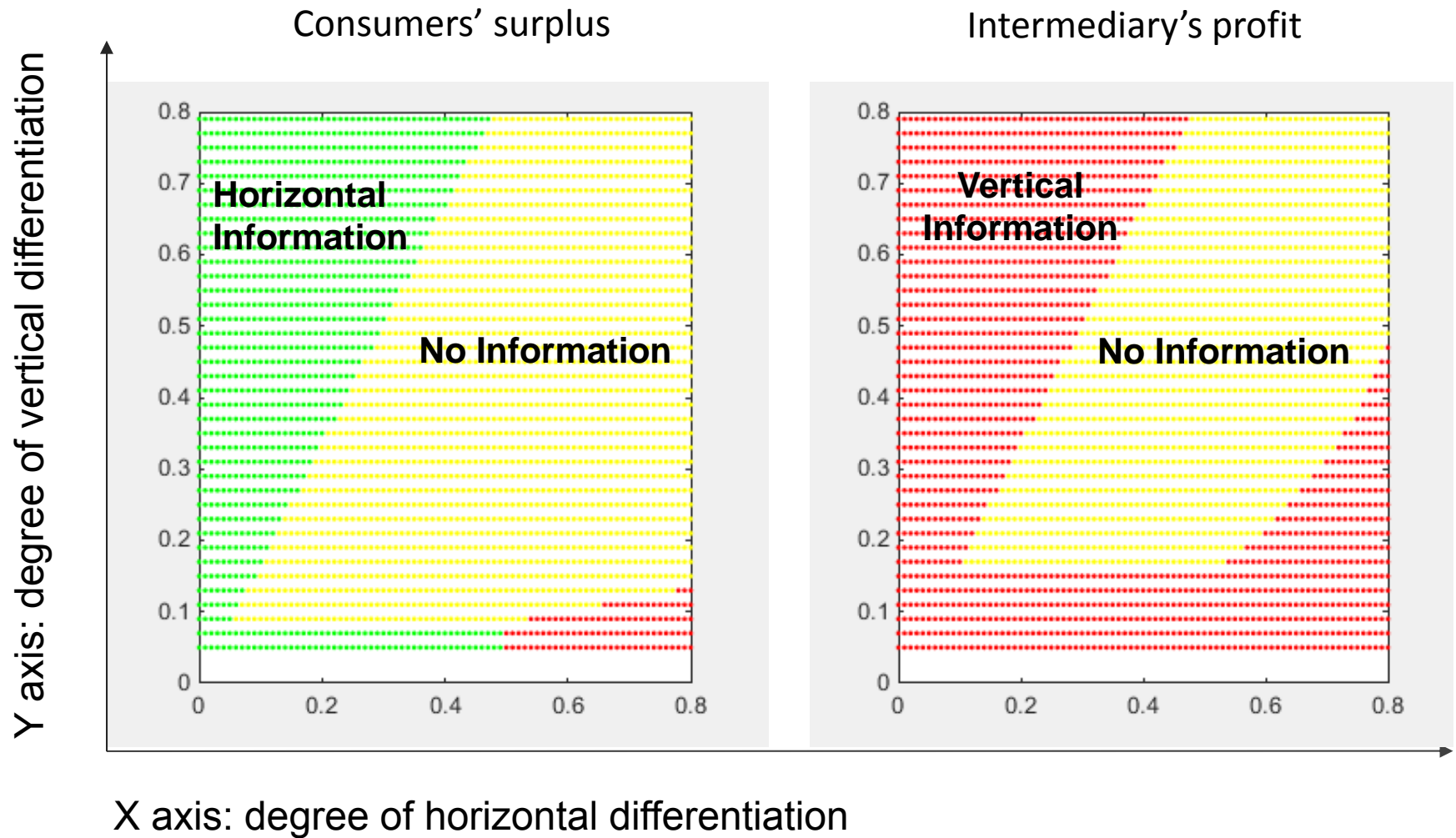
→ **Equilibrium Concept:** Nash Equilibrium for Second-Price Auctions

2. Through simulations of the model, we analyze how the outcome in terms of consumers' welfare, intermediary's profit and firms' profit changes under the different scenarios

Welfare Analysis



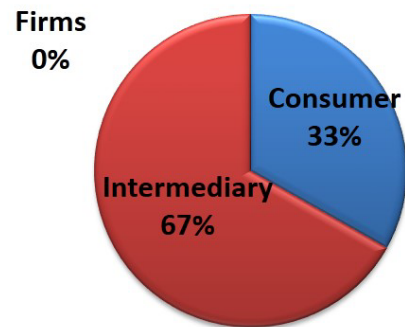
Welfare Analysis



Welfare Analysis

- Allocation of Benefits (proportions) among Consumer, Advertiser and Intermediary under the four scenarios

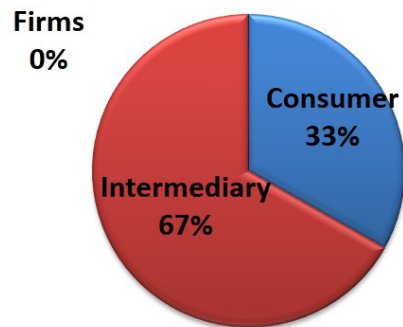
a. No Information



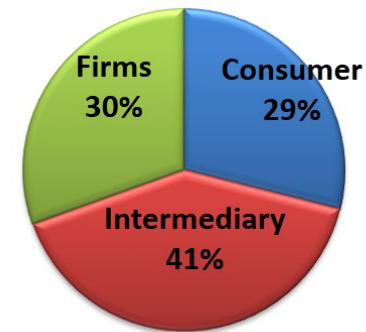
Welfare Analysis

- Allocation of Benefits (proportions) among Consumer, Advertiser and Intermediary under the four scenarios

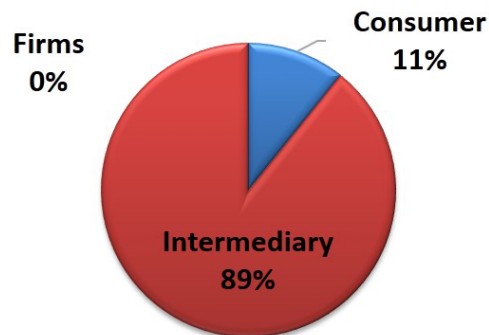
a. No Information



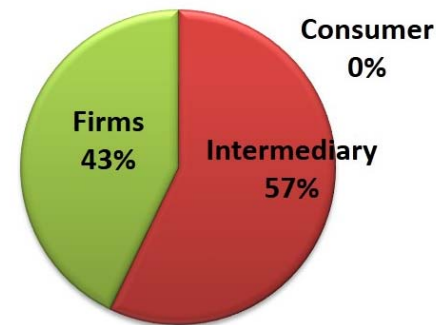
b. Horizontal Information



c. Vertical Information



d. Complete Information




Results

1. Consumer's surplus is higher when only specific type of information is available (horizontal information) and, generally, when less information is available
2. There exist situations in which the incentives of the Intermediary are misaligned with respect to consumer's interest
3. Under certain conditions, the Intermediary obtains the highest proportion of benefits from the targeting process
4. A strategic intermediary may choose to selectively share consumer data in order to maximize its profits

Limitations/Future work

- Competition among ad networks
- Costs/investments for ad networks
- Reduction in consumer search costs
- Empirical analysis



Personal information is the
lifeblood of the Internet

Loss of privacy is the price to pay
for the benefits of big data

Sharing personal data
is an economic win-win

How is the surplus generated by
personal data allocated?

Who bears the costs of privacy
enhancing technologies?

When do consumers benefit
from trades in their data?

*“The Economics of Privacy,” Acquisti, Taylor, and Wagman,
Journal of Economic Literature, (forthcoming)*

Catherine Tucker

Massachusetts Institute of Technology

Privacy Protection, Personalized Medicine and Genetic Testing

Co-author: Amalia Miller (University of Virginia)



**Privacy Protection,
Personalized
Medicine and
Genetic Testing**

Amalia R. Miller and
Catherine Tucker

Our research question

What kinds of privacy
protections encourage or
discourage the spread of
hospital genetic testing
for cancer?

What can genetic tests be used for?

Identifying genetic information to predict

- susceptibility to disease
- course of disease
- response to treatment.

Pr

Angelina's

DOUBLE
MASTECTOMY

Inside

BRCA1 mutation

New Details
THE OHIO THREE



HOW THEY
SURVIVED

Look at state law variation from 2000-2010 which echoes 3 approaches to privacy

- Informed consent (EU privacy directive of 1996?)
- Regulating data use (US approach?)
- Establishing property rights over data (Coasian)

We use a national survey to understand who gets a genetic test

- National Health Interview Surveys (NHIS) - part of CDC
- In 2000, 2005, 2010 they asked 30k survey takers about genetic testing.

There are pros and cons of the dependent variable

- . Yes: Testing for predictors of breast, ovarian cancer.

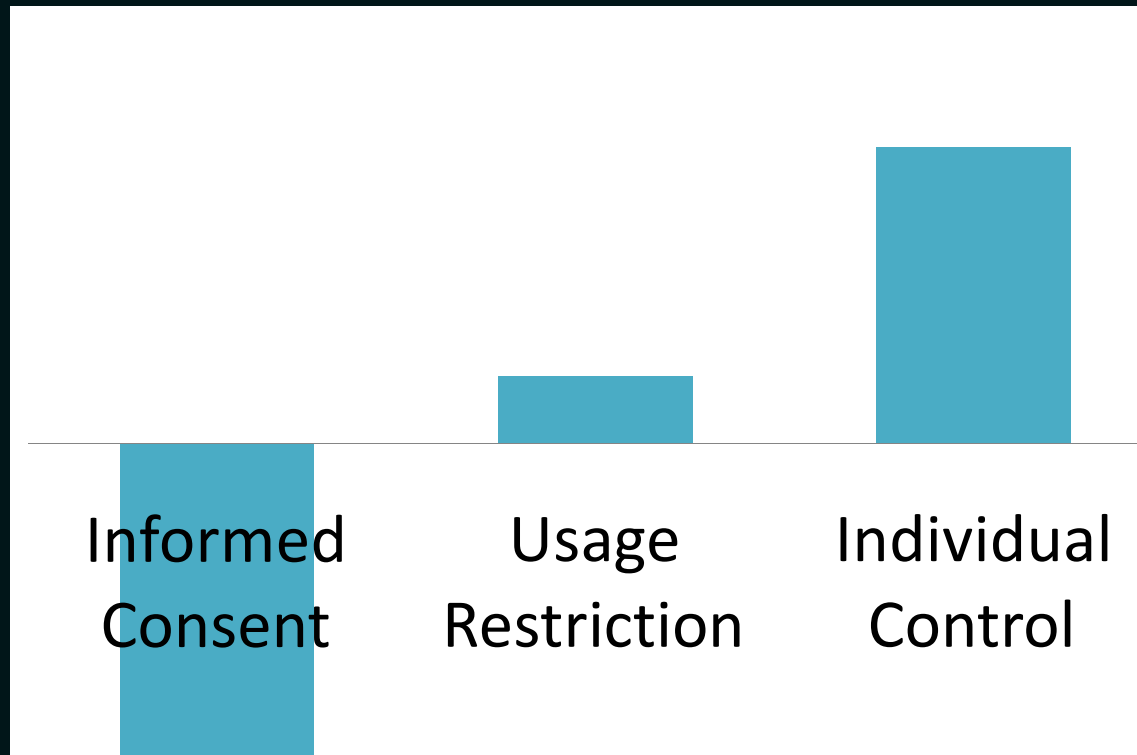
Actionable.

- . But: Few positive observations ($< 1\%$)

We use standard econometric techniques

- Statistically relate the decision to take a genetic test to changes in the patient's state's privacy law.
- See the paper for the equations and methods

Informed Consent reduces genetic testing by one third, Individual Control increases genetic testing by one third



The controls had weak but expected effects

- Female, black, family cancer positively affect decision
- No insurance (weakly) negatively affects decision
- State characteristics, age, private insurance aren't significant

The positive effect for Individual Control is not driven by hospitals

- . Hospitals react negatively to consent laws
- . But also react negatively to patient property rights

We provide evidence that our effect is causal with placebos

- . No effect for genetic laws for HIV testing - not driven by tastes for privacy
- . No effect for genetic laws on flu shots - not driven by tastes for preventative care

What is going on?

- Discrimination laws - lack of information?
- Consent without control - highlights powerlessness?
- Data ownership - Perception of control or Coase?

Pure consent

I understand that DNA analysis may yield information on biological paternity, the results of which will not be disclosed to me unless biological paternity is relevant to the reason for which I have submitted this DNA sample. In addition, I agree to provide a family history that will be complete and correct to the best of my knowledge. I further understand that that genetic counseling may be important for me (my child) depending on the results of this testing and I have been informed and have been provided with information identifying genetic counselors should I wish to consult with one. I understand that the procedure used to collect the blood or tissue samples has inherent, but minimal risks that have been explained to me, and that additional blood or tissue sample(s) may need to be obtained if the results of the original testing are inconclusive. I understand that my (my child's) DNA will be stored in the repository maintained by PerkinElmer Genetics in Pittsburgh, PA or at its responsible delegated institution or repository.

By my signature below, I hereby consent to PerkinElmer Genetics, and its responsible delegated institutions or repositories, to use and disclose my individually identifiable medical information (including without limitation all associated genetic information) for purposes of my diagnosis and/or treatment (e.g., to my (my child's) treatment providers), to seek payment from third parties for such testing and to conduct ongoing health care operations (e.g., administrative oversight, quality assurance/control and technical innovations).

I understand that to the extent any technical innovation or invention is developed by PerkinElmer Genetics or its responsible delegated institutions or repositories in connection with the testing, quality control or other permitted use of my (my child's) blood or tissue sample, neither I nor my child shall be entitled to any compensation with respect thereto. Your signature on this form indicates that you understand the information regarding molecular genetic testing and agree to obtain such testing. In no way does this waive your legal rights or release PerkinElmer Genetics or its responsible delegated institutions and repositories from their legal and professional responsibilities. If you have further questions concerning matters related to this consent, please discuss them with your medical geneticist, genetic counselor, or referring physician.

Revised 3/26/08

(Signature of patient or legal guardian)

Description of authority (parent, guardian, etc.)

Date

Signature of Witness

Consent with Property Rights

I understand that:

1. A blood sample or other type of specimen will be obtained using a procedure which carries a very slight risk of bleeding or infection.
2. The test results will be communicated to me by my physician and/or genetic counselor in a confidential manner and will not be released to another party without my signed consent unless required by law. Results will become part of my Medical Genetics record.

Our effects appear to be driven by privacy concerns

- Larger effects for those with higher underlying risk
- No effects for those with past cancer diagnosis
- Larger effects for 'privacy-protecting' individuals

Summing Up

There are of course limitations

1. The unobserved
2. No information about interpretation
3. Early stage of diffusion.

When states give more control over how their private information is shared genetic testing increases

- . Particularly for those who are more worried about 'bad news'
- . Hospitals respond negatively

We find that informed consent
deters patients and hospitals
from testing

Data usage policies have little effect

- . Good or bad news depending on how you look at it

Thank you! æetucker@mit.edu

Sasha Romanosky

RAND Corporation

Examining the Costs and Causes of Cyber Incidents



Costs and Consequences of Cyber Incidents

Sasha Romanosky



RAND

Institute for Civil Justice

Motivation

- Data breaches and privacy violations have become commonplace, affecting thousands of firms, and millions of individuals,
 - yet we don't fully understand their costs or impacts
 - nor do we properly understand the firm's incentives to invest in cyber security controls
- Therefore, using a dataset of 12,000 events, we examine the costs, scale, and overall risk of events, by industry and over time

Four types of cyber events

Data breaches

Unauthorized disclosure of personal info

Security incidents

Computer attacks against a company

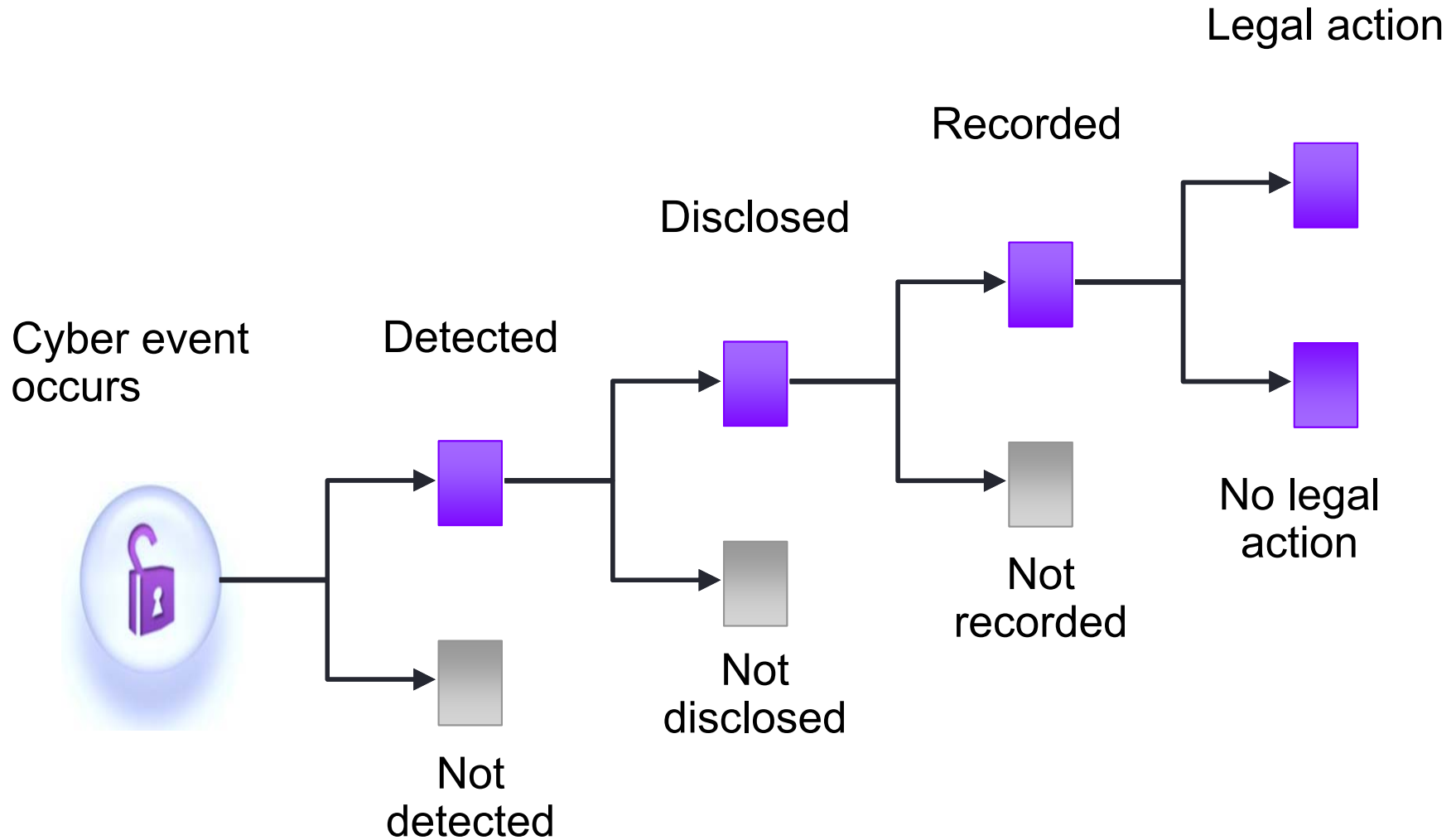
Privacy violations

A company's willful collection or use of personal info

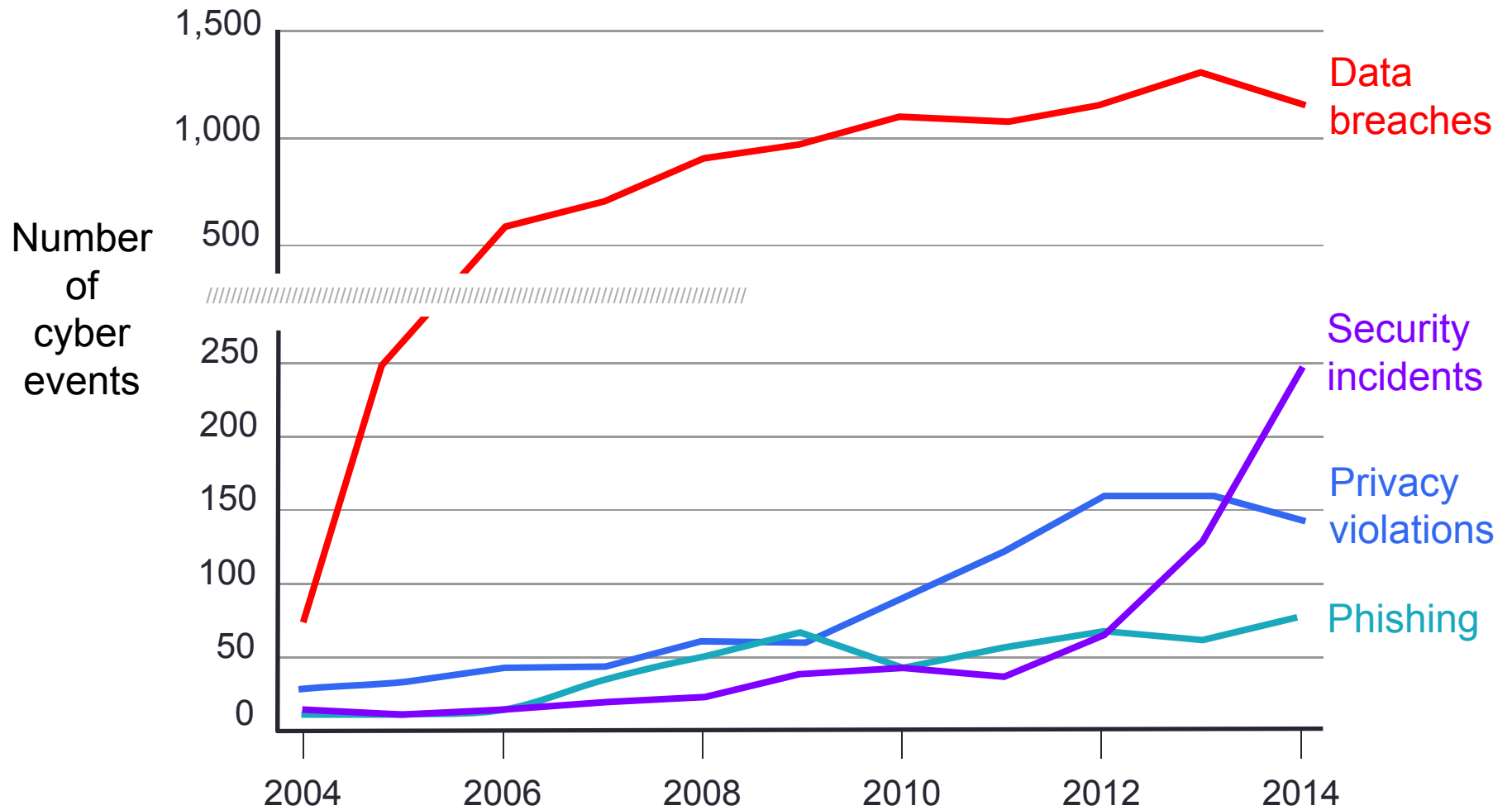
Phishing/skimming

Financial criminal acts committed against individuals and firms

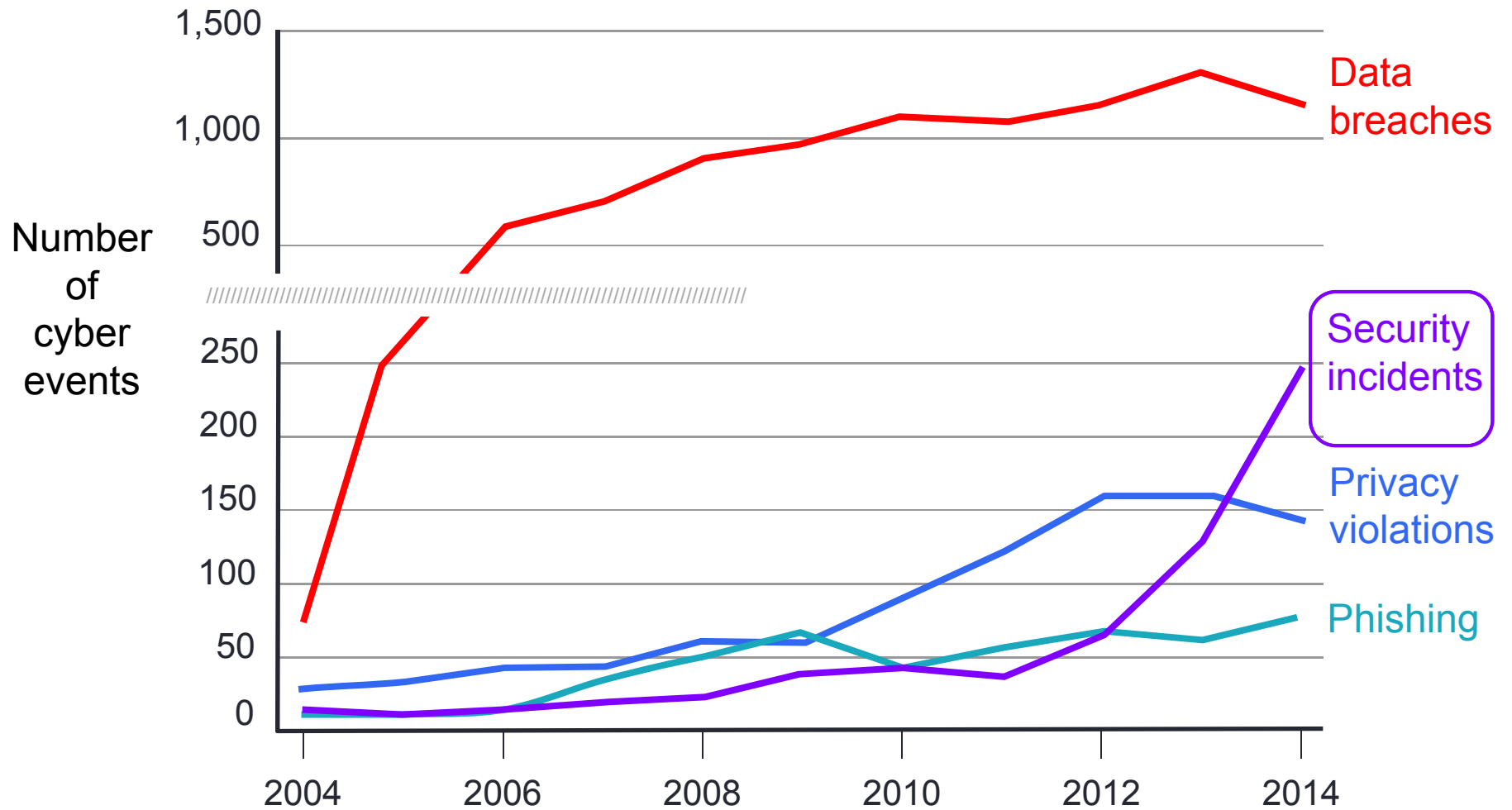
We observe publicly available data



Data breaches greatly outnumber all other incidents



But security incidents are increasing rapidly

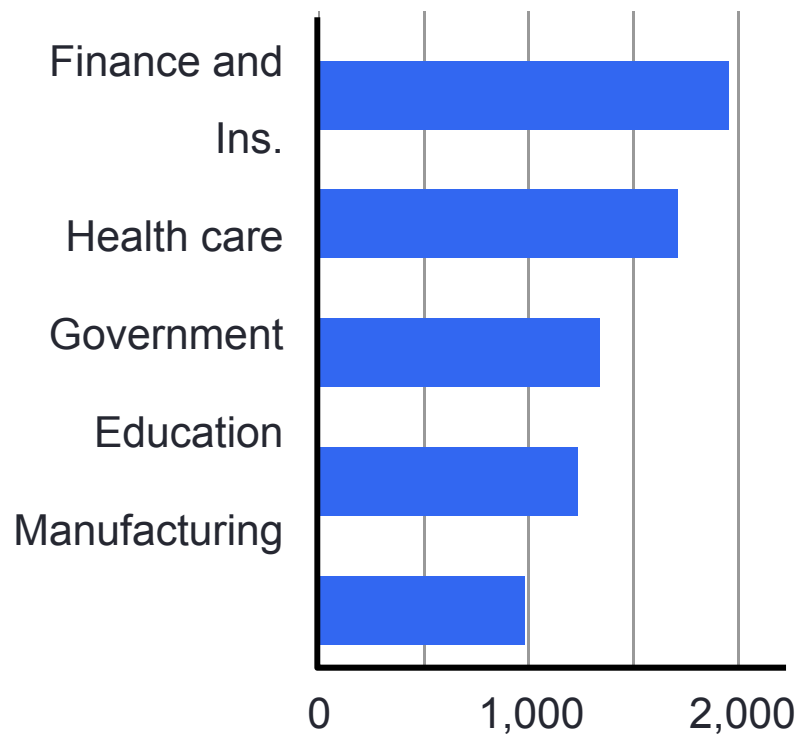


Industry analysis

- There are many ways to understand risk by industry:
 - Total incidents, and incident rate
 - Total lawsuits, and litigation rate
 - Costs per event
- This helps us understand which industries pose the greatest risk

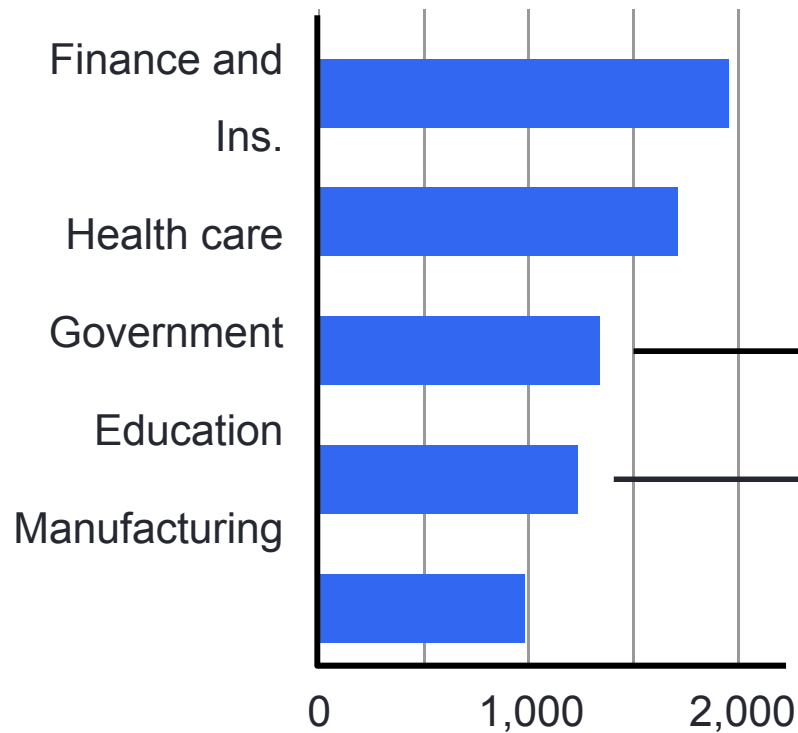
Finance and Insurance, and Health Care sectors suffer highest number of incidents

Total incidents

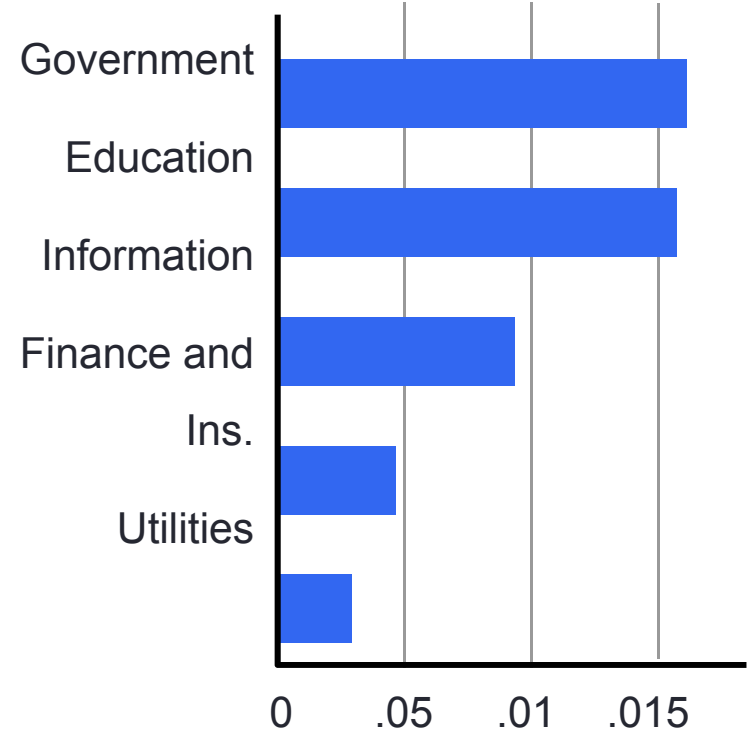


But Govt, and Education sectors suffer highest incident rates

Total incidents

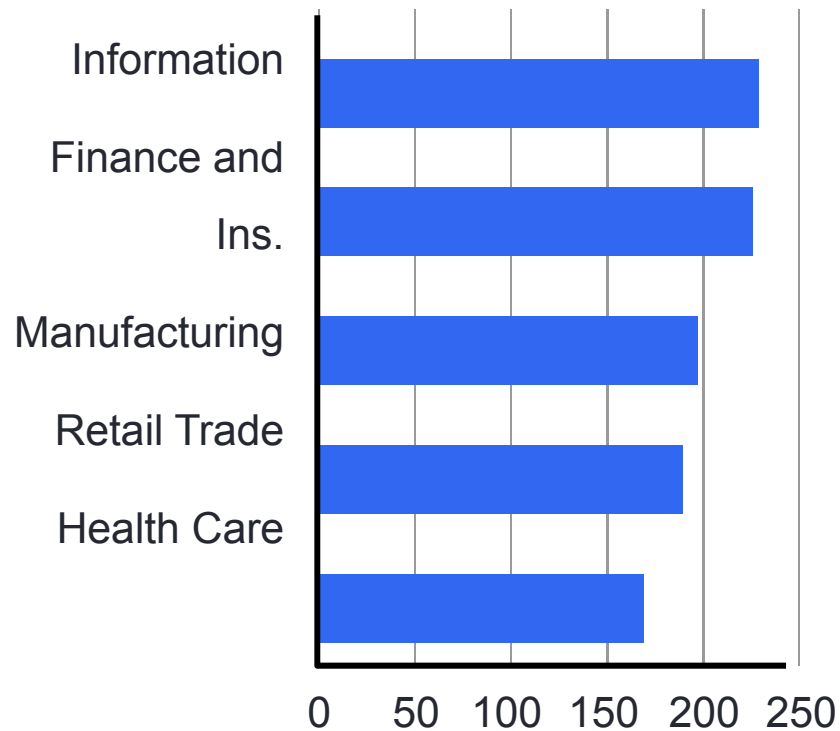


Incident rate



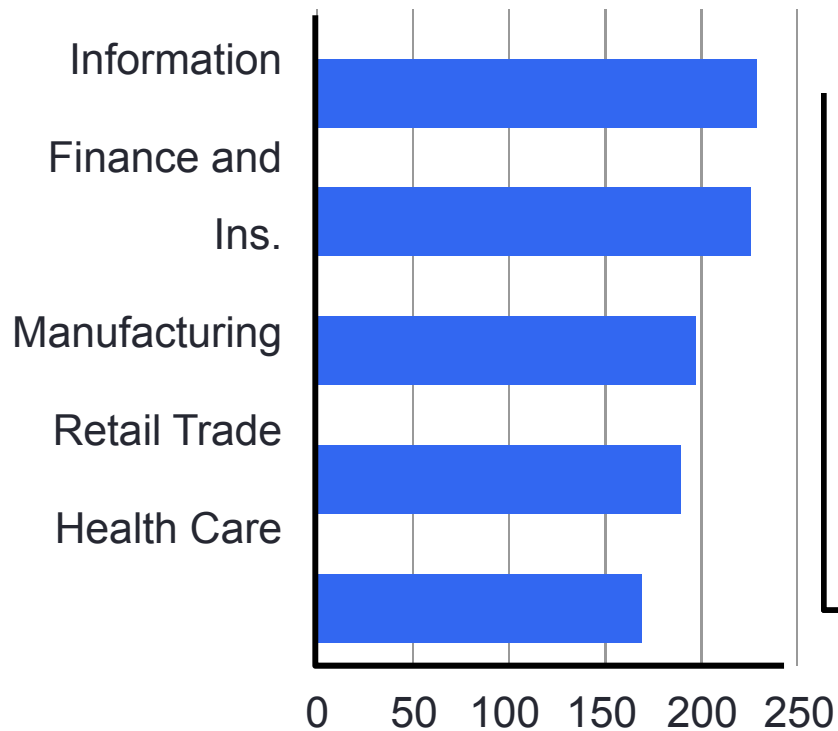
Firms in Information and Finance/Insurance sectors are most often litigated

Number of lawsuits

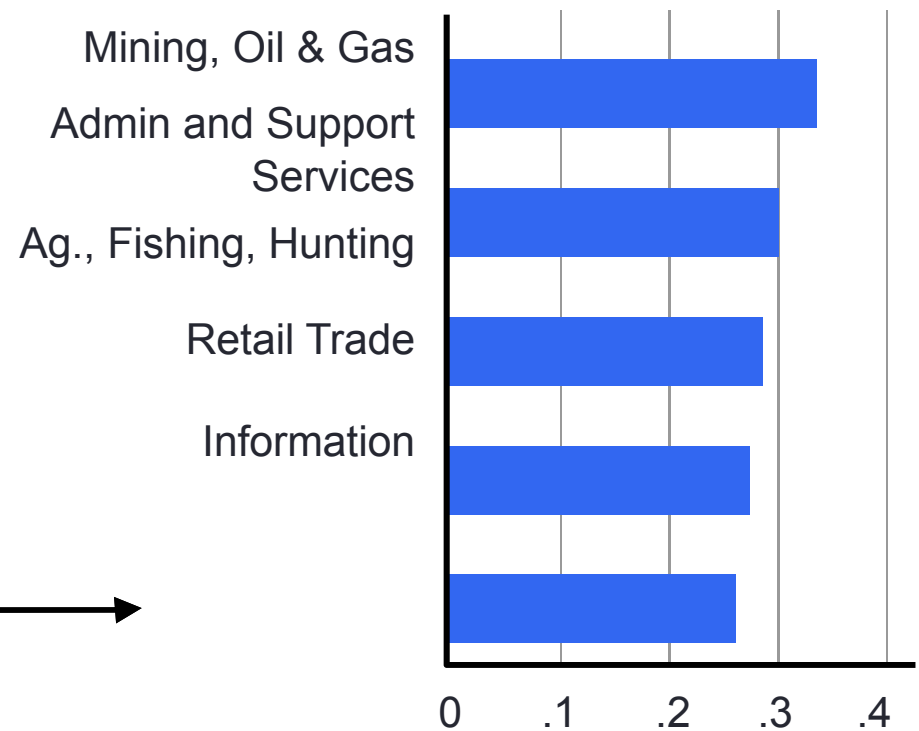


But, Oil & Gas suffers the highest litigation rate

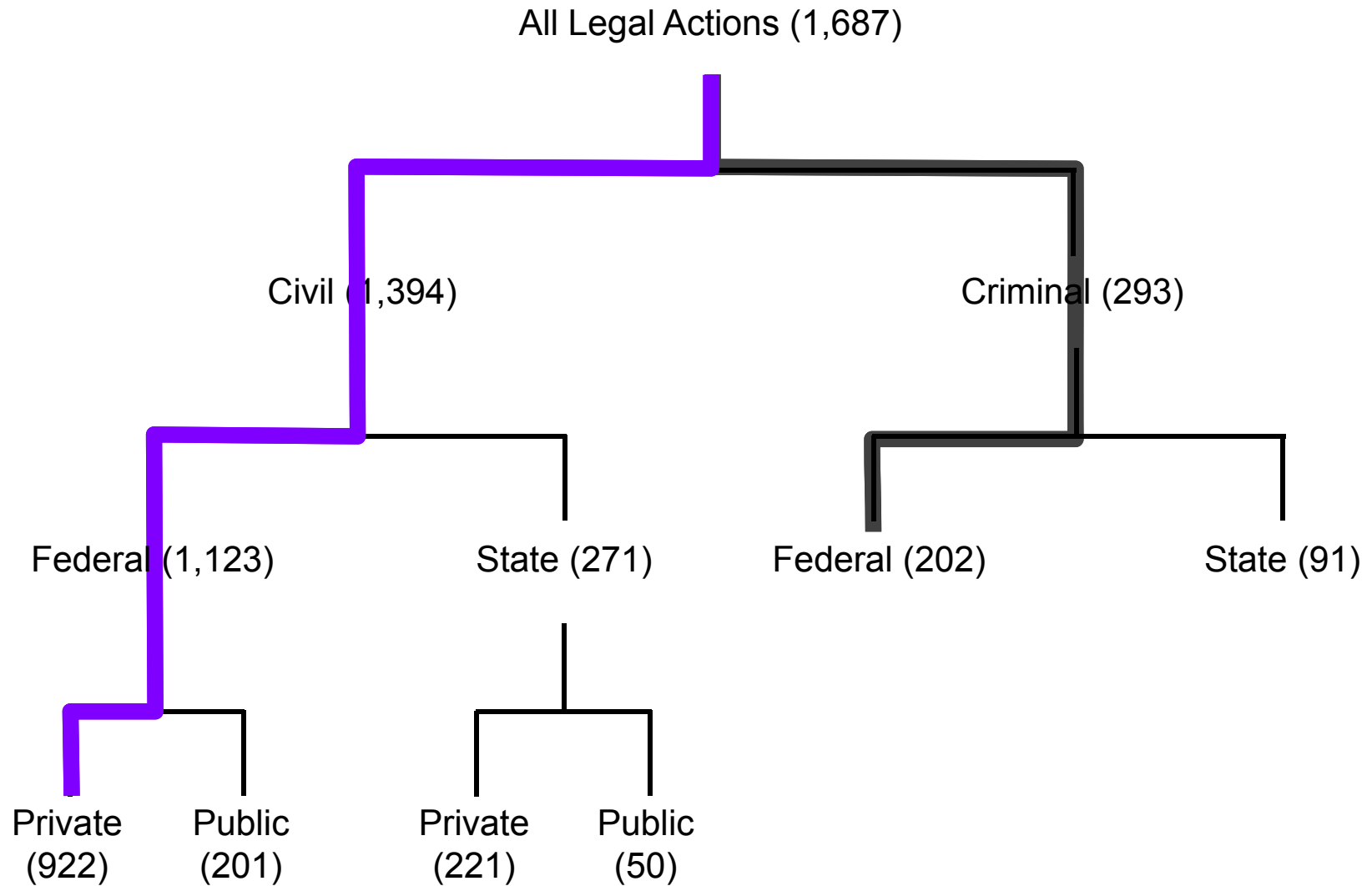
Number of lawsuits



Litigation rate

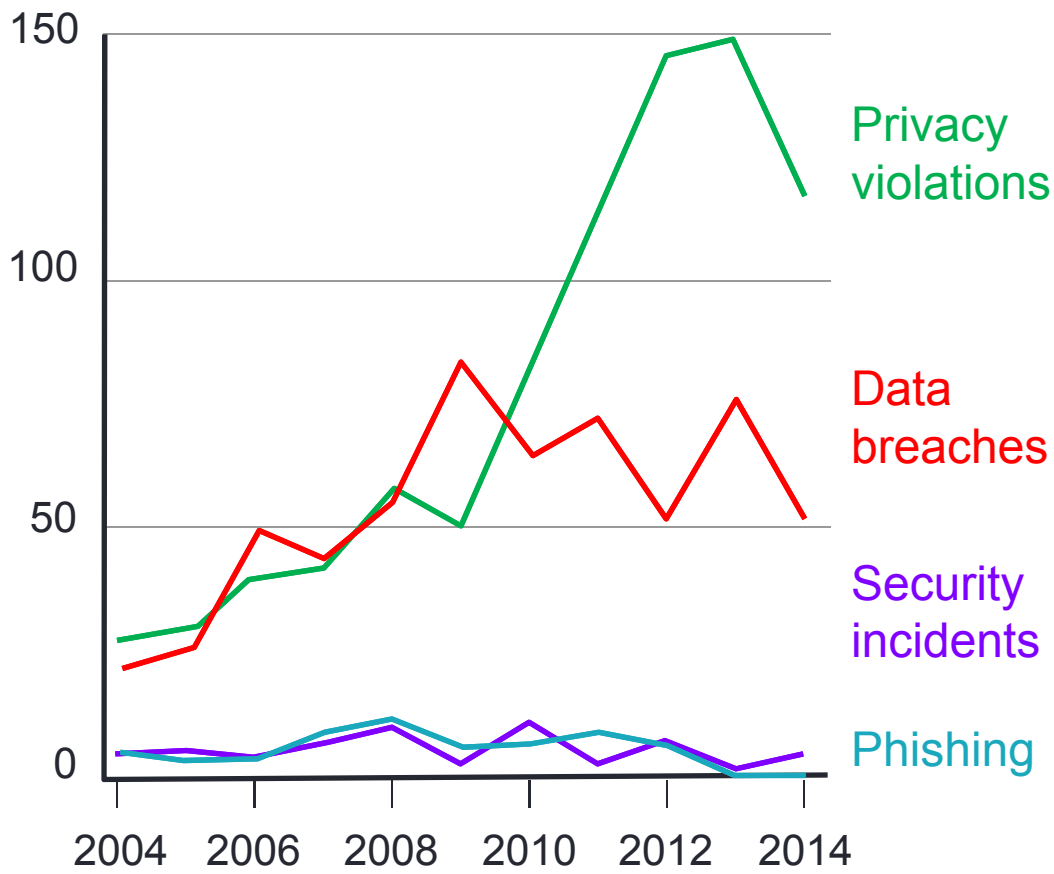


Next, let's examine legal actions



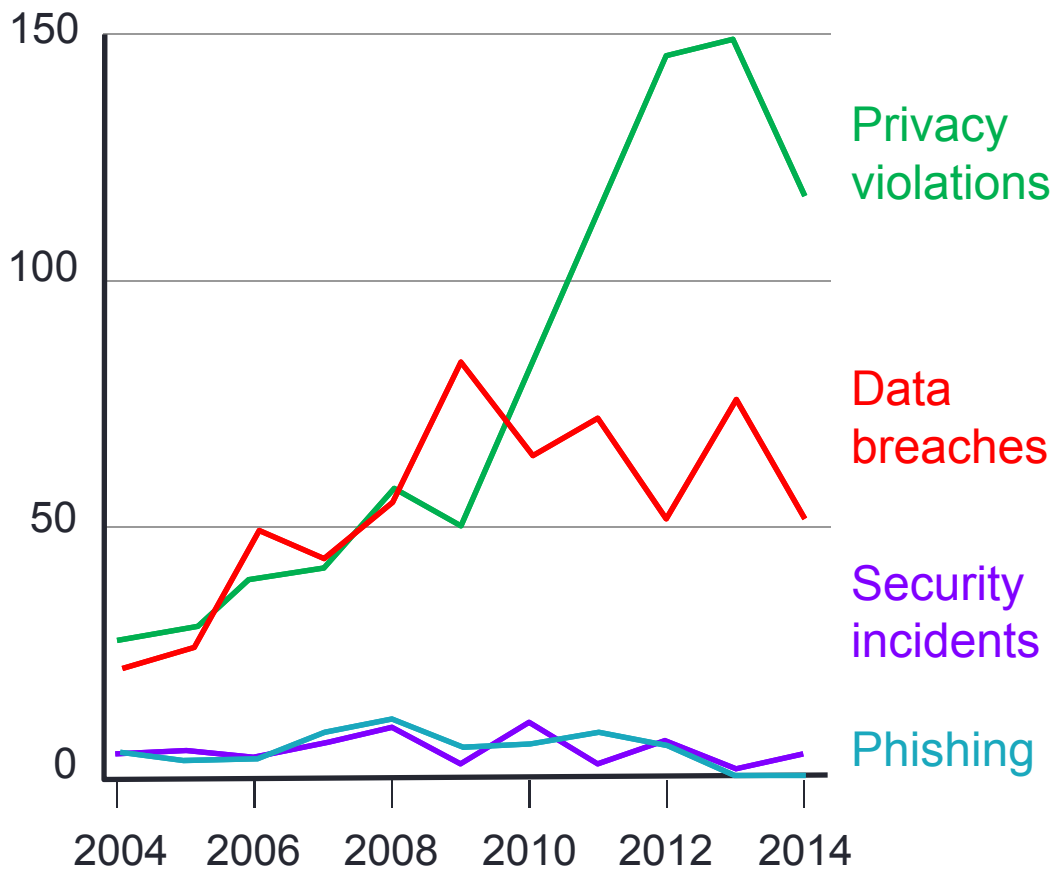
Privacy litigation has increased sharply

Total number of lawsuits

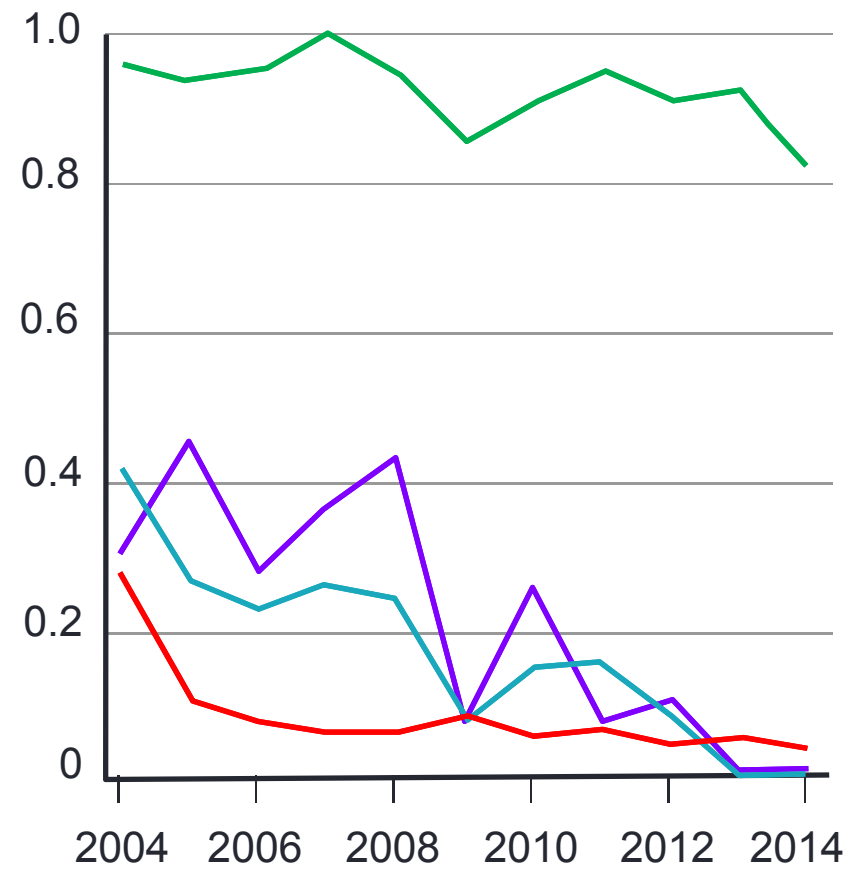


But overall litigation rates are declining

Total number of lawsuits



Litigation rate



Most data breaches cost firms less than \$200K

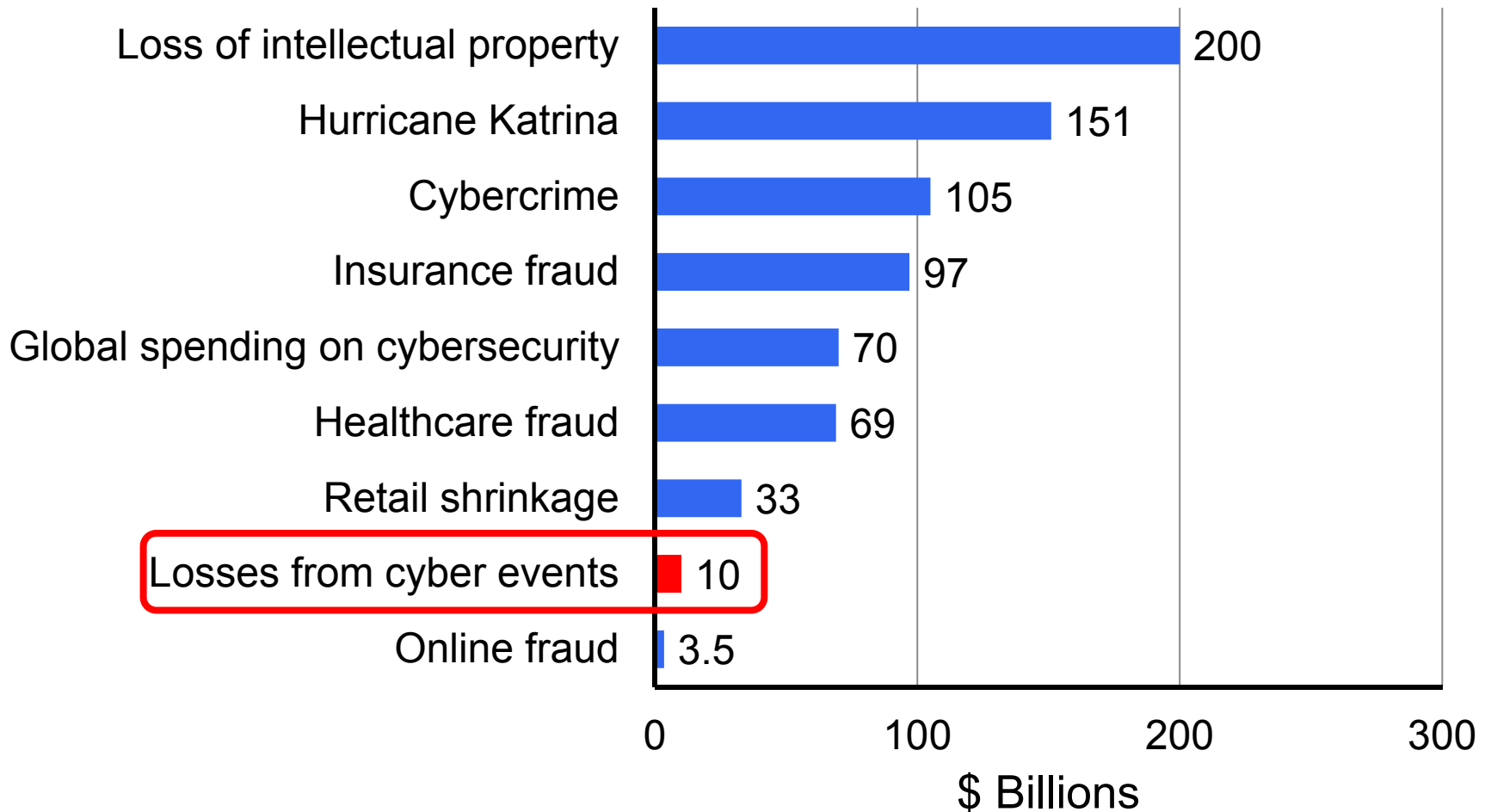
	Min	Max	Median	N
Data Breaches	\$25	\$572m	\$170 K	602
Security Incidents	\$100	\$100m	\$330 K	36
Privacy Violations	\$180	\$750m	\$1.34 M	234
Phishing/Skimmi ng	\$0	\$710 m	\$150 K	49

These costs are much lower than the \$5m often cited 963

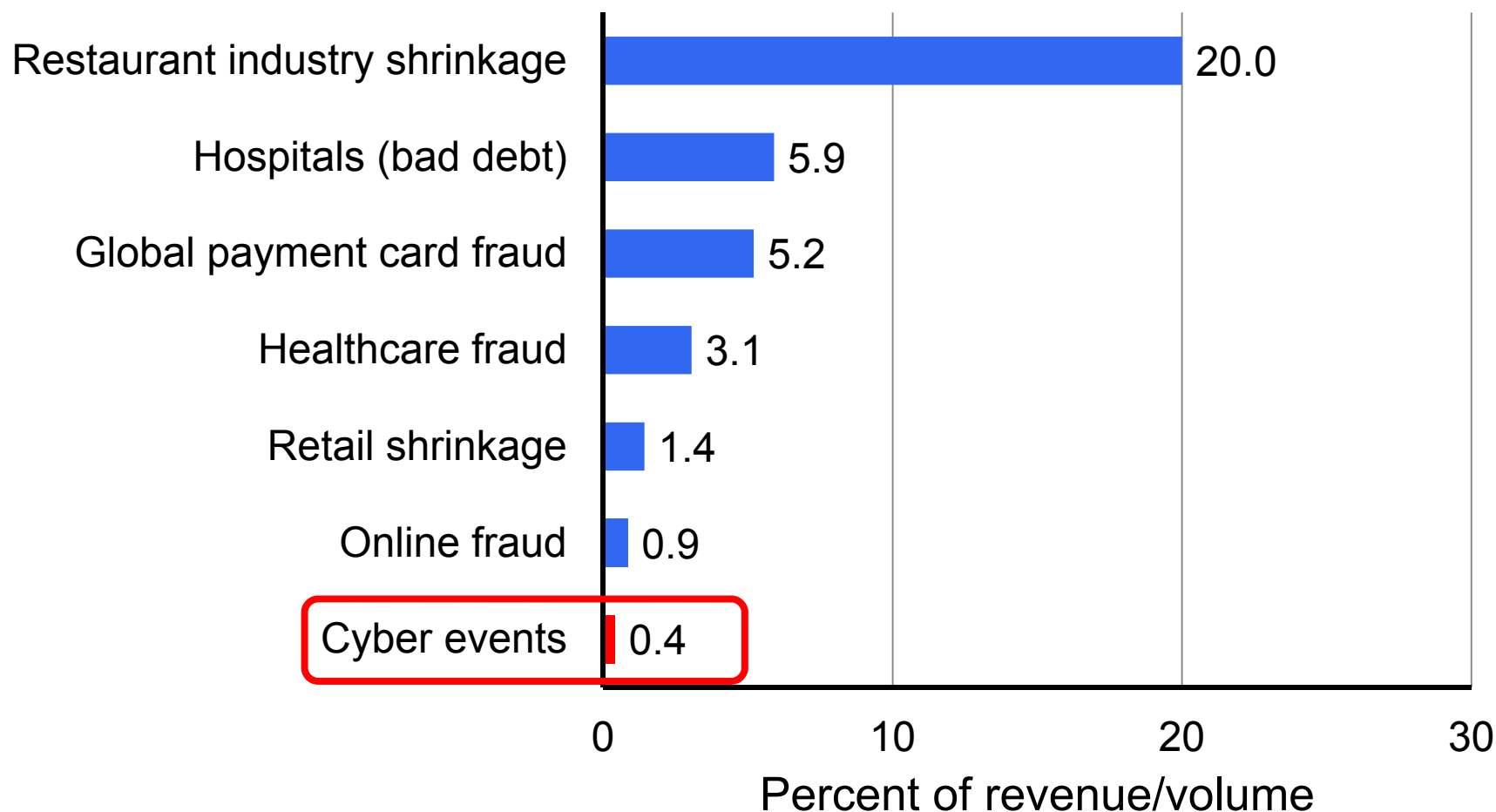
Repeat Players

- 38% of firms (almost 4800) in our dataset suffered multiple incidents
- 50% of all incidents within the Information and Finance/Insurance industries involve repeat players
- No significant difference in legal actions or litigation rate for this group, relative to single players
- However, data breach costs are twice as large for repeat players:
 - \$9.8m vs \$4m for single players

Annual losses from cyber events are comparatively small



As a percent of revenue,
the cost of cyber events is also very small



Troubling paradox; where do the incentives lie?

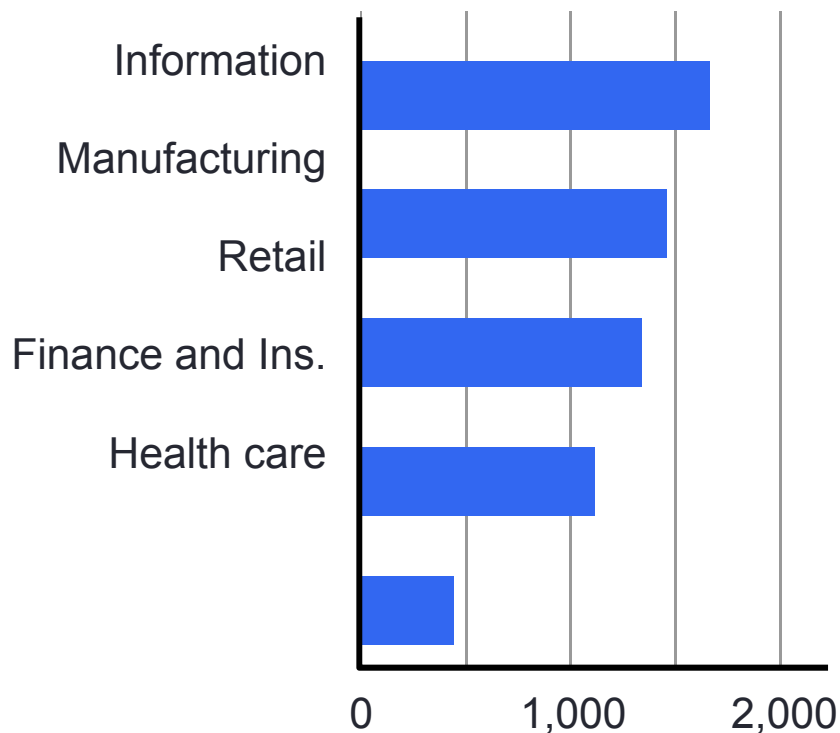
- On one hand, cyber events *and* legal actions are increasing
 - Compromising the most sensitive kinds of personal information
- On the other hand, typical costs are relatively small
 - And consumers seem quite satisfied with firm responses (Ablon et al, 2016)
- What does this suggest for firm incentives in cyber security?

Questions?

sromanos@rand.org

Retail, Information, and Manufacturing sectors suffer highest losses

Total losses (in Millions \$)



Loss per event



Losses are typically of two types

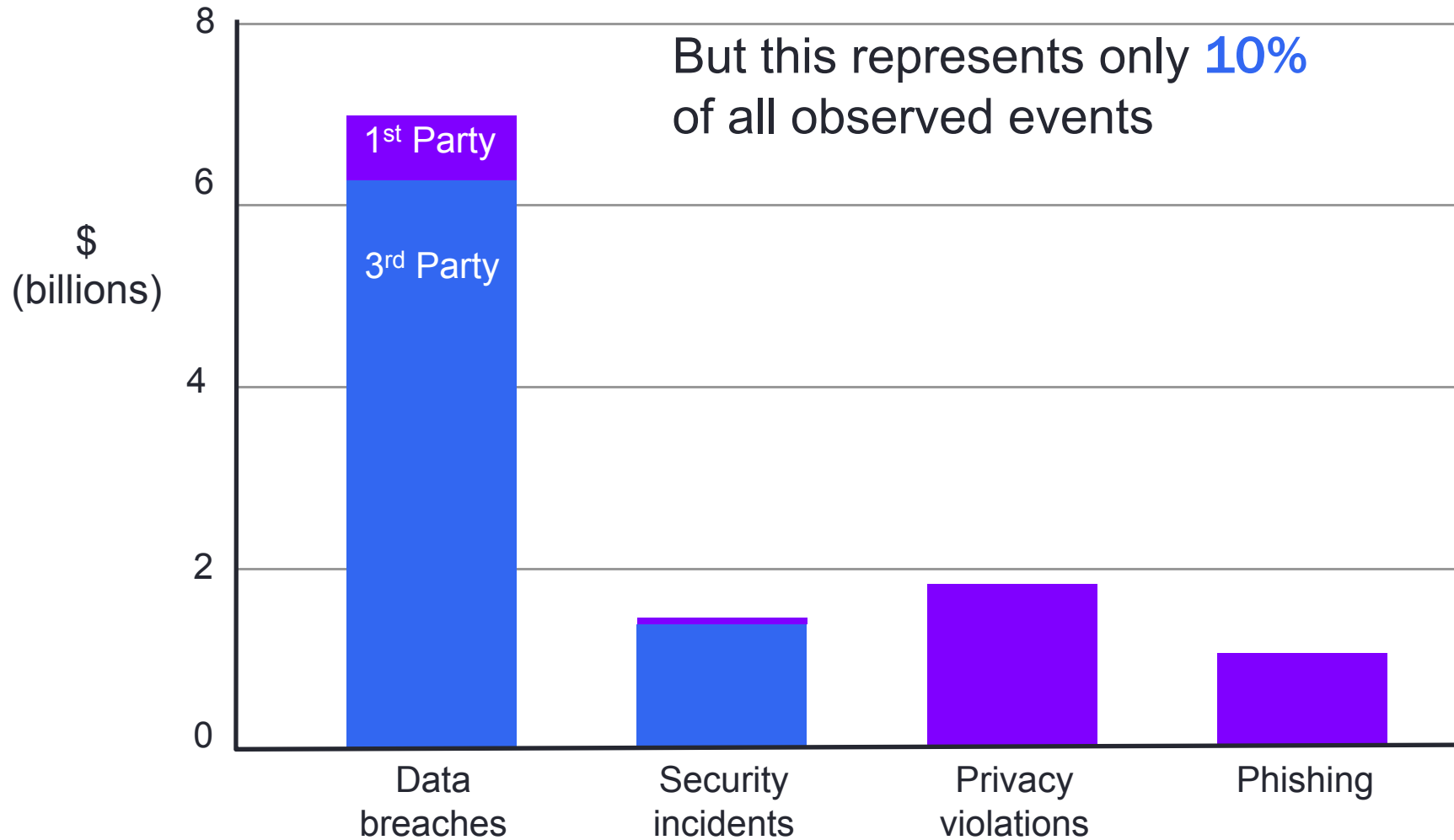
1st-party losses

- Breach notification, counsel, forensics, IT repair, consumer redress
- Also includes money stolen from banks, financial companies

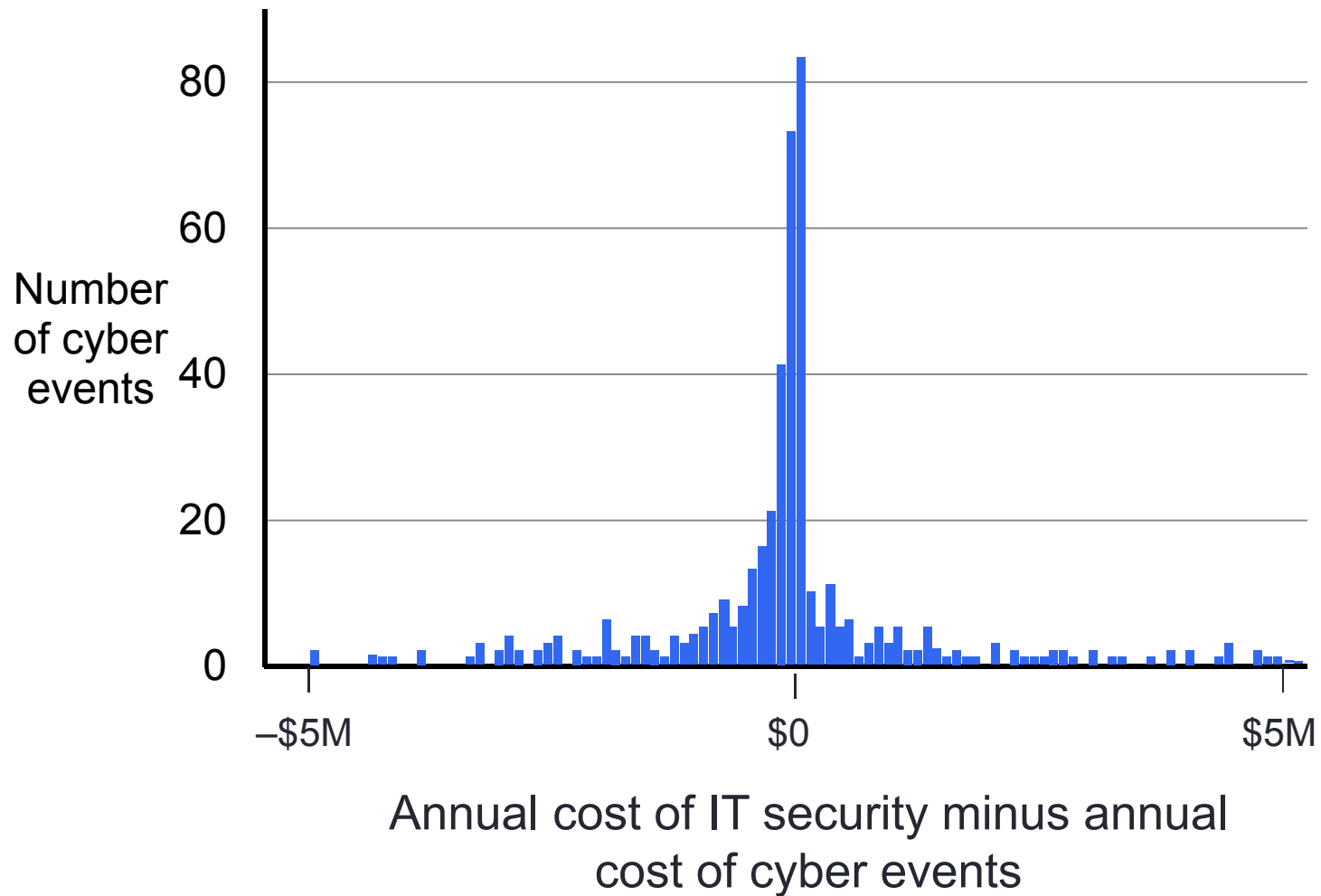
- Settlements and other judicial awards
- Administrative rulings, *cy pres*

3rd-party losses

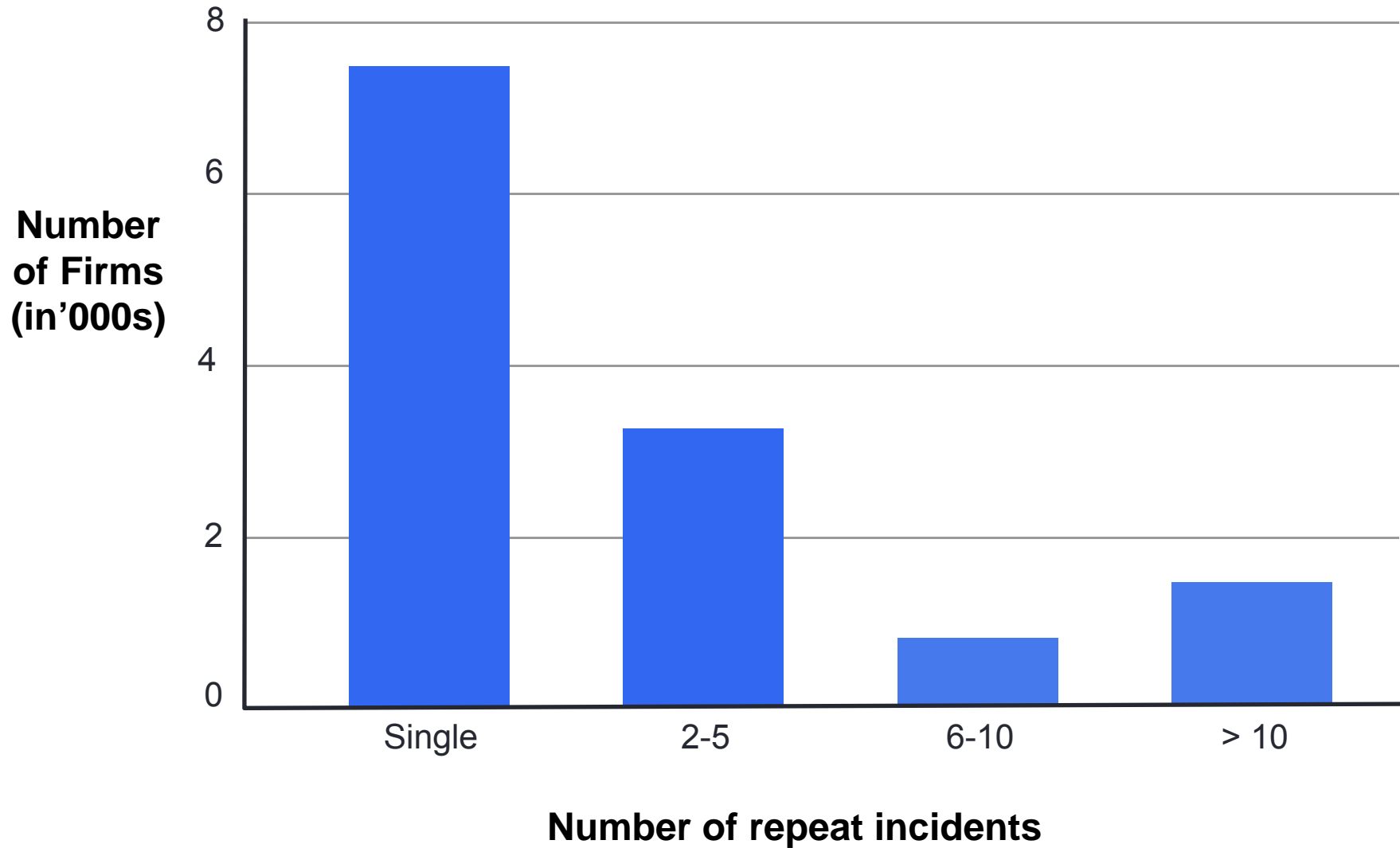
Total recorded losses of \$10 billion



In most cases, cyber events cost firms about what they spend on IT security



Distribution of Repeat Players



Discussion of Session 4

Discussants:

- **Kevin Moriarty**, Federal Trade Commission
- **Doug Smith**, Federal Trade Commission
- **Siona Listokin**, George Mason University

Presenters:

- **Jens Grossklags**, Pennsylvania State University
- **Veronica Marotta**, Carnegie Mellon University & **Alessandro Acquisti**, Carnegie Mellon University
- **Catherine Tucker**, Massachusetts Institute of Technology
- **Sasha Romanosky**, RAND Corporation

