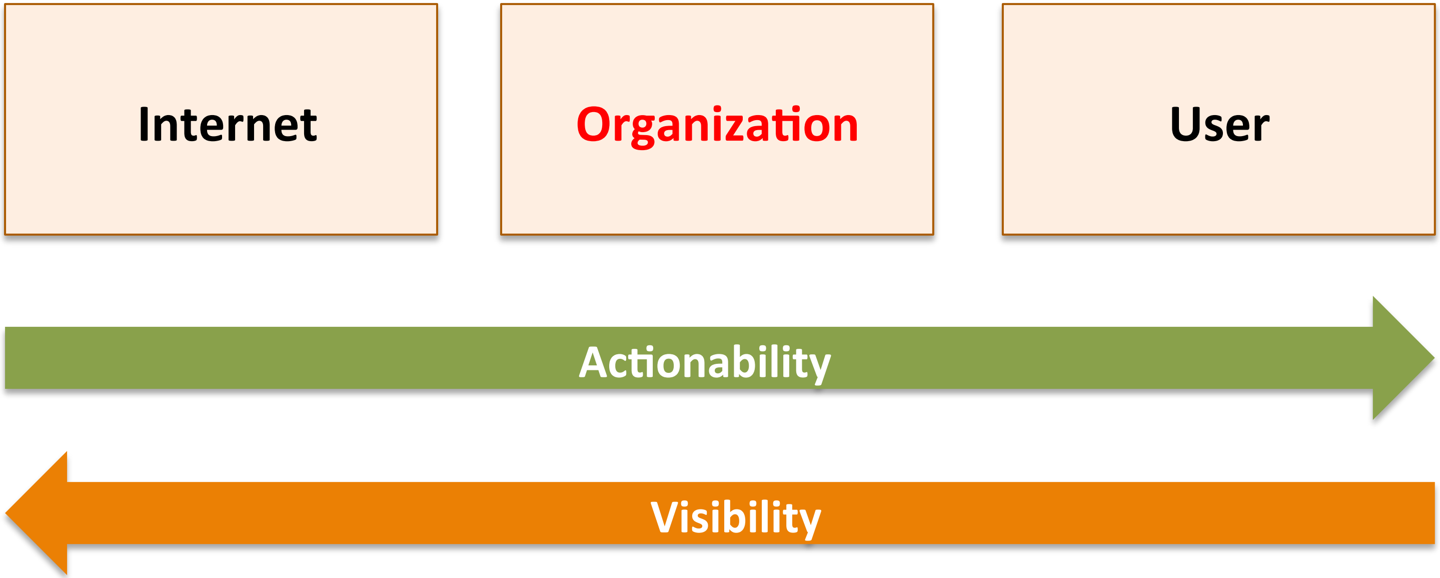
Learning from Early Attempts to Measure Information Security Performance

Jing Zhang[1], Robin Berthier[2], Will Rhee[1], Michael Bailey[1],
Partha Pal[3], Farnam Jahanian[1], and William H. Sanders[2]

[1] University of Michigan
[2] University of Illinois at Urbana-Champaign
[3] BBN Technologies, Cambridge, MA

Importance of Organizations in the Security Ecosystem



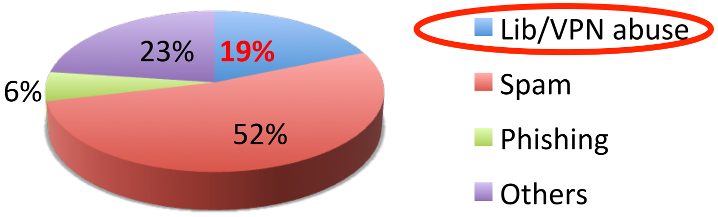
Our Organizations

- Security operation teams at our universities
  - Information and Infrastructure Assurance (IIA) at University of Michigan
  - Security teams at UIUC
- We oversee IT security at the universities
  - More than 40,000 students
  - More than 30,000 faculty and staff at UofM, and more than 10,000 at UIUC
  - Facilitating campus-wide incident response activities
  - Provide services such as security assessment and consultation, network scans, education and training
  - Managing IT security issues at the university level

Organizational Background

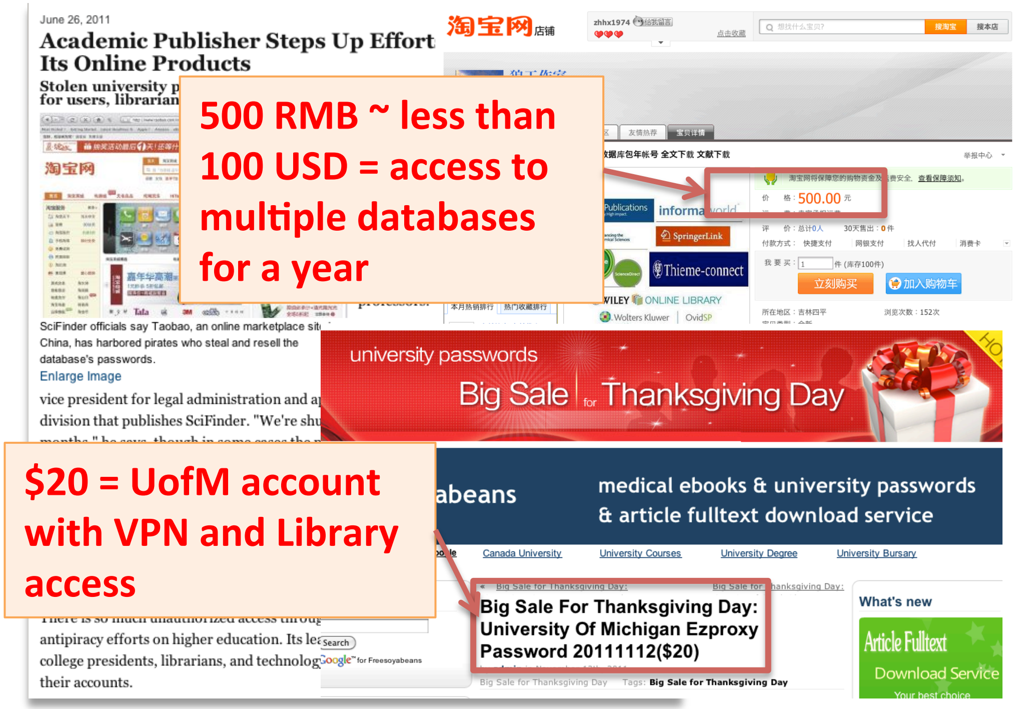
Context: Account compromise at UofM and UIUC

- **613 incidents** related to unauthorized use of university accounts during 2010 and first 6 months of 2011 at UofM



- **178 compromised accounts** were reported in the first half of 2011 at UIUC

- Market place for the compromised university accounts



June 26, 2011

Academic Publisher Steps Up Effort
Its Online Products

Stolen university p
for users, librarian

SciFinder officials say Taobao, an online marketplace sit
China, has harbored pirates who steal and resell the
database's passwords.

Enlarge Image

vice president for legal administration and a
division that publishes SciFinder. "We're shu

There is so much unauthorized access throug
antipiracy efforts on higher education. Its lea
college presidents, librarians, and technolog
their accounts.

数据库包年帐号 全文下载 文献下载

淘宝网将保障您的购物资金及 费安全。查看保障须知。

价 格：500.00 元

评 价：总计0人 30天售出：0 件

付款方式： 快捷支付 网银支付 找人代付 消费卡

我 要 买：1 件 (库存100件)

所在地区：吉林四平 浏览次数：152次

university passwords

Big Sale for Thanksgiving Day

medical ebooks & university passwords
& article fulltext download service

Big Sale For Thanksgiving Day:
University Of Michigan Ezproxy
Password 20111112($20)

Big Sale for Thanksgiving Day Tags: Big Sale for Thanksgiving Day

500 RMB ~ less than 100 USD = access to multiple databases for a year

$20 = UofM account with VPN and Library access

Organizational Goals

- We want to answer:
  - How secure is the organization?
  - Has the secure posture improved over the last years?
  - How to compare with peers with respect to security?
  - What is the marginal change in the security, given the use of a tool or practice?
  - How to prioritize resources to maximize security and minimize risks?
- Security Metrics
  - Micro-level of view
  - Quantified measurement
  - Hard to achieve
    - Complexity of the environment
    - rapid evolution of technology and adversarial action

"We cannot manage what we cannot measure!"

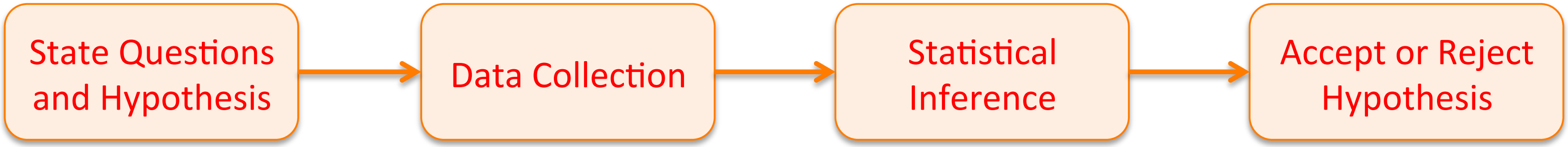
Our Work Today

What we have

- Incidents Tickets
- Authentication Logs
- Victim Information
- Password-cracking results
- Security quiz results

Factors Analyzed

- Victim Demographic
- Temporal Factor
- Geographical Factor
- Topological Factor
- Usage Behavior
- Password Strength
- Security Quiz



State Questions and Hypothesis → Data Collection → Statistical Inference → Accept or Reject Hypothesis

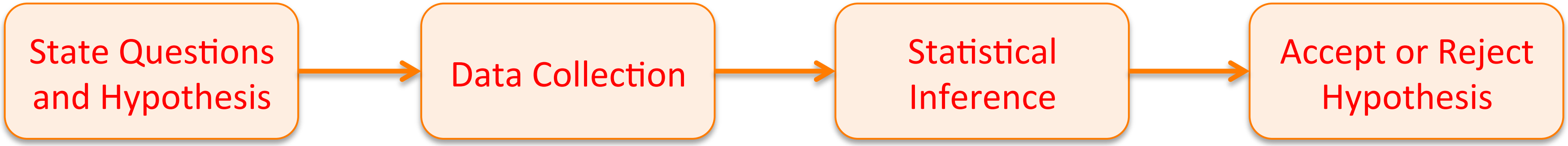
Our Work Today

What we have

- Incidents Tickets
- Authentication Logs
- Victim Information
- Password-cracking results
- Security quiz results

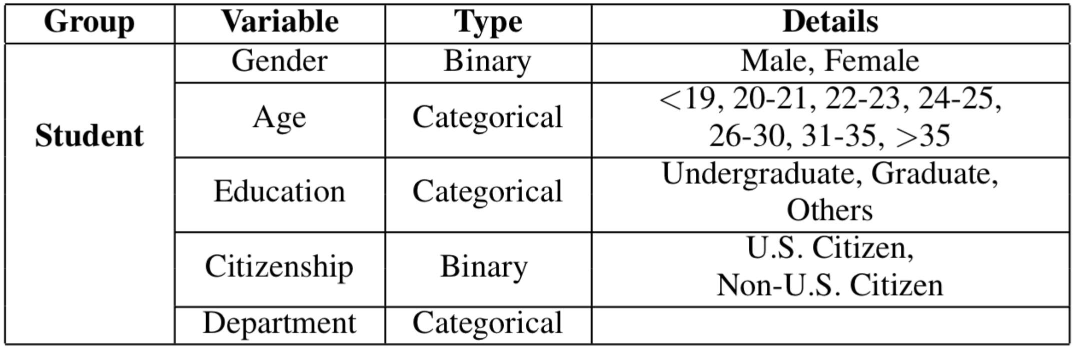
Factors Analyzed

- Victim Demographic
- Temporal Factor
- Geographical Factor
- Topological Factor
- Usage Behavior
- Password Strength
- Security Quiz



Example 1 - User Susceptibility

- Question
  - What roles *gender, age, education-level, citizenship,* and *department* play in the compromise of student accounts?
- Data
  - Student victims: *242* at UofM from 2009 to 2011, *130* at UIUC in 2011
  - Aggregated Demographics for the total student population



Group	Variable	Type	Details
Student	Gender	Binary	Male, Female
	Age	Categorical	<19, 20-21, 22-23, 24-25, 26-30, 31-35, >35
	Education	Categorical	Undergraduate, Graduate, Others
	Citizenship	Binary	U.S. Citizen, Non-U.S. Citizen
	Department	Categorical	

Example 1 - User Susceptibility

- Methodology: Multivariate Linear Regression
  - Predict the effect of one factor, *holding other factors constant*.
  - Example: Age and Education Level
    - Simple distribution -> 20-21, undergraduate
    - Undergraduate students has more people in age 20-21 than graduate students
    - Which is the real significant factor? Or both?
  - Logistic Regression Model:

$$L = a + \sum B_i X_i. \qquad L = \ln \frac{\hat{p}}{1 - \hat{p}}.$$

    - Null Hypothesis Ho: Bi = 0 (Variable Xi is not statistically significant in predicting user susceptibility)
    - Test Statistics: p-value < 0.05

Example 1 - User Susceptibility

- Results

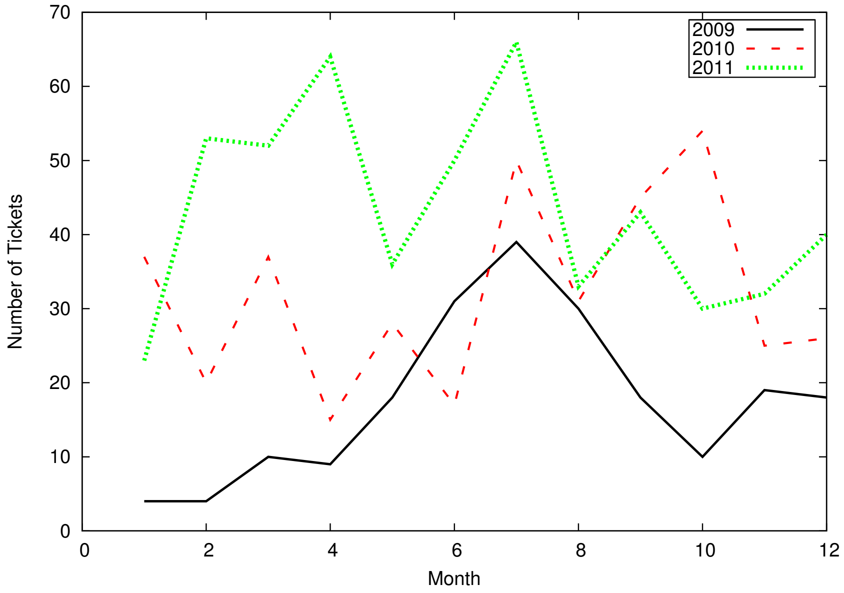
Factor	University	Year	p-value	Coef.
Undergraduate	UofM	2009	0.009	2.957
		2010	<0.001	3.520
		2011	0.020	3.489
	UIUC	2011	0.958	-10.733
Age (20-21)	UofM	2009	0.002	1.219
		2010	0.004	0.823
		2011	0.017	0.896
	UIUC	2011	0.410	-0.472
Citizenship	UofM	2009	0.520	0.315
		2010	0.659	-0.126
		2011	0.128	-0.460
	UIUC	2011	0.007	0.5433

Disagreement between the two universities

Gender is not statistically significant!

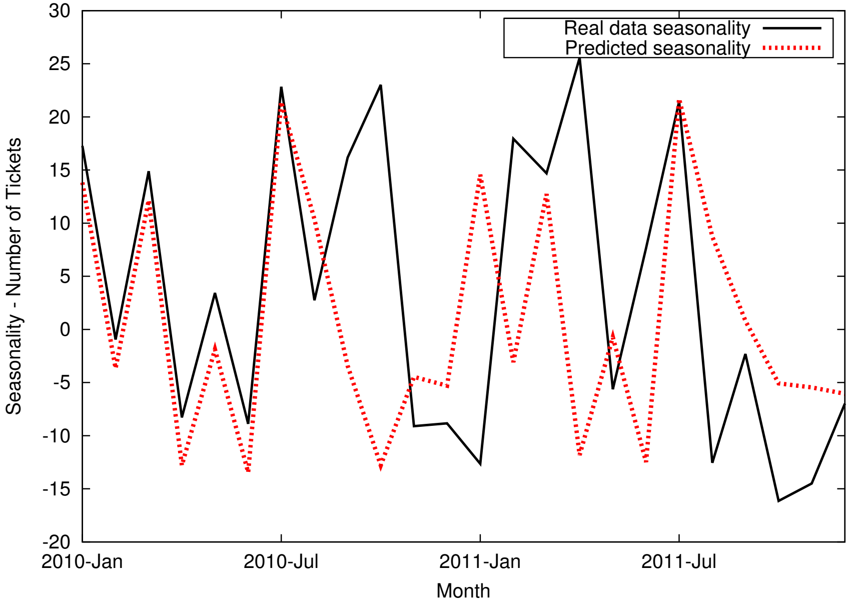
Example 2 - Temporal Factor

- Question
  - Whether the incidence of compromises varies at different time of the year?
- Data
  - Monthly number of tickets at UofM from 2009 to 2011



Example 2 - Temporal Factor

- Methodology: Time series data analysis
  - "Holt-Winters" exponential smoothing procedure
    - Long-term trend + Seasonality
- Result
  - No seasonality pattern in the monthly number of tickets



Creation time ≠ Compromise time

Example 3 - Password Policy

- Question
  - Whether accounts with weak passwords are more likely to be compromised?
- Data
  - Password-cracking performed at UofM (2012)

	# of total	# of compromised	Pr (compromise)
Weak Password	2,284	12	0.525%
Total Population	550,000	380	0.069%

Example 3 - Password Policy

- Methodology: Test of Homogeneity
  - Whether the response of identifiable sub-populations differ from those of others
  - Null Hypothesis H_o: users who have weak passwords have the same probability to be compromised as other users
  - Test statistics: deviance; Confidence level: p-value < 0.05
- Result
  - Test statistics of deviance of 28.09 and a p-value of 1.16^{-16}
  - Reject Null Hypothesis, and conclude that the users, who use weak passwords, have a higher probability to be compromised

Is weak password the reason of compromise?

But are the limited number of potentially impacted accounts worth our effort?

Discussions

- Are the questions meaningful?

Actionable? Proactive or Reactive?

- Is it the right data?

Quality? Sensitivity?

- Are we using right analysis techniques?

Observation ≠ Statistical Inference
Correlation ≠ Causality

- How to reproduce the measurement?
  - Continuous measurement
  - Reproduce across multiple organizations

Generalized measurement metrics and techniques
Data collection and sharing platform

- How to form actionable strategies based on those metrics?

Results ≠ Strategy
Strategy ≠ Success