The Strategy of Cyber: Ambiguity, Bias, Economics, Obscurity, and Structure

Chris Hazard
Hazardous Software Inc.
Hazardous Software: AI, Strategy, & Serious Games
Outline

- Motivation
- Cyber Threat Defender (Hazardous Software)
- DACDAM (NATO)
- Game Theory & Economics
  - INMAST (Hazardous Software / NCSU LAS)
- Information Theory & Ambiguity
- Trust & Psychology
Cyber: A Game of Resources

- Increased vulnerability
  - Global and open culture
  - Network effect of communication scale
  - Target rich environment

- Increased threat
  - Existing infrastructure built without enough security
  - Easier to acquire resources
  - Attribution is harder and less predictable
Data Breaches

- Healthcare: 95% hacking/malware/virus
- Retail: 75% due to insiders selling information
- Education: 50% due to stolen paperwork
- Finance: 70% of employee-related breaches were due to selling information to outsiders

Data Breaches

- Healthcare: 95% hacking/malware/virus
- Retail: 75% due to insiders selling information
- Education: 50% due to stolen paperwork
- Finance: 70% of employee-related breaches were due to selling information to outsiders

Data Breaches

- Healthcare: 78% from theft of electronic devices
- Retail: Targeted by outsiders, 50% involving paper records
- Education: 75% hacking/malware/virus
- Finance: 67% hacking/malware/virus, 71% of employee-related breaches were accidental

People like to be validated and like to be right
Humans Are Rational*
(and so are adversaries)

*given limited computational bounds, heuristics, unfounded beliefs of others, inaccurate capability assessments, inexplicable valuations, and some level of [im]patience
Cyber Threat Defender
Multicriteria Planning in Cyber

● Goals:
  ● Keep internal & external customers satisfied
  ● Reduce successes of attacks
  ● Reduce adverse impact of attacks
  ● Stay within budget

● Constraints: integrity, availability, confidentiality, resiliency, redundancy, graceful degradation, performance, regulations, physical attack surface, virtual attack surface, human factors
Use Cases

● Experiment with offensive & defensive strategies
● Organizational knowledge transfer
● Communicate and evaluate decisions, strategies, architectures, and attacks
● Analyze risks
● High-level control for cyber ranges or live systems (given sufficiently automated provisioning & configuration)
● Training and familiarization with cyber domain
● Better understanding of infrastructure
  ● IBM zSystems has deployed it as a sales tool
Attacks Against IT Dept

• Decrease Customer Satisfaction w/ Service
  • Increase system load
  • Disable components
  • Destroy data
  • Destroy components

• Harm Business Resources
  • Reduce number of satisfied customers
  • Force budget reduction

• Ransom
Attacker Verb Examples

- Exfiltrate
- Exploit Exfiled Data
- Acquire Internal Resource
- Attempt to Acquire Attack Vector
- DDoS
- Audit
- Flood
- Power Outage
Resource Property Examples

Installation Cost
Main’t Cost
Transaction Service Rate and Characteristics
Transaction Latency Characteristics
Vulnerability Discovery Rate
What things can a resource investment in cyber defence impact?
How do we measure the impact of a cyberattack?
What things are not under the control of the organization?
What things are under the control of the organization?
How do attackers interact with one another?
How do system administrators interact with one another?
How much freedom is in modifying the policies of the organization?
NATO’s Data-Farmable Agent-based Cyber Defence Assessment Model (DACDAM)
Verizon 2016 Data Breach Report
## Cordon & Search
Counter Insurgency (COIN) Patrols

<table>
<thead>
<tr>
<th>Term</th>
<th>Contrast</th>
<th>Length-t Ratio</th>
<th>Individual p-Value</th>
<th>Simultaneous p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>number-of-attackers</td>
<td>0.014499</td>
<td>4.33</td>
<td>0.0010*</td>
<td>0.0010*</td>
</tr>
<tr>
<td>sensor-p-detect</td>
<td>0.006166</td>
<td>1.85</td>
<td>0.0766</td>
<td>0.7637</td>
</tr>
<tr>
<td>shut-down-threshold</td>
<td>-0.005368</td>
<td>-1.61</td>
<td>0.1133</td>
<td>0.9096</td>
</tr>
<tr>
<td>avg-replication-on-servers</td>
<td>0.005136</td>
<td>1.54</td>
<td>0.1270</td>
<td>0.9963</td>
</tr>
<tr>
<td>mean-vulnerabilities-learned</td>
<td>0.004831</td>
<td>1.45</td>
<td>0.1490</td>
<td>0.9637</td>
</tr>
<tr>
<td>hacker-learning-time</td>
<td>-0.003616</td>
<td>-1.06</td>
<td>0.2656</td>
<td>0.9996</td>
</tr>
<tr>
<td>pc-vulnerabilities</td>
<td>0.003597</td>
<td>1.08</td>
<td>0.2682</td>
<td>0.9997</td>
</tr>
<tr>
<td>shut-subnet-threshold</td>
<td>0.002157</td>
<td>0.65</td>
<td>0.5350</td>
<td>1.0000</td>
</tr>
<tr>
<td>mean-vulnerabilities-known</td>
<td>-0.001278</td>
<td>-0.38</td>
<td>0.7142</td>
<td>1.0000</td>
</tr>
<tr>
<td>susceptibility-to-phishing</td>
<td>0.001085</td>
<td>0.31</td>
<td>0.7676</td>
<td>1.0000</td>
</tr>
<tr>
<td>pc-percent-vulnerabilities</td>
<td>0.000769</td>
<td>0.23</td>
<td>0.8253</td>
<td>1.0000</td>
</tr>
<tr>
<td>max-competency</td>
<td>-0.000686</td>
<td>-0.21</td>
<td>0.8448</td>
<td>1.0000</td>
</tr>
<tr>
<td>mean-time-to-update</td>
<td>0.000664</td>
<td>0.20</td>
<td>0.8504</td>
<td>1.0000</td>
</tr>
<tr>
<td>min-competency</td>
<td>0.000079</td>
<td>0.02</td>
<td>0.9831</td>
<td>1.0000</td>
</tr>
<tr>
<td>number-of-attackers*number-of-attackers</td>
<td>-0.002237</td>
<td>-0.67</td>
<td>0.5047</td>
<td>1.0000</td>
</tr>
<tr>
<td>number-of-attackers*sensor-p-detect</td>
<td>0.006277</td>
<td>1.83</td>
<td>0.0720</td>
<td>0.7474</td>
</tr>
<tr>
<td>sensor-p-detect*sensor-p-detect</td>
<td>0.002223</td>
<td>0.67</td>
<td>0.5159</td>
<td>1.0000</td>
</tr>
<tr>
<td>number-of-attackers*shut-down-threshold</td>
<td>-0.004973</td>
<td>-1.40</td>
<td>0.1383</td>
<td>0.9534</td>
</tr>
<tr>
<td>sensor-p-detect*shut-down-threshold</td>
<td>-0.004543</td>
<td>-1.36</td>
<td>0.1722</td>
<td>0.9829</td>
</tr>
<tr>
<td>shut-down-threshold*shut-down-threshold</td>
<td>-0.000798</td>
<td>-0.24</td>
<td>0.8198</td>
<td>1.0000</td>
</tr>
<tr>
<td>number-of-attackers*avg-replication-on-servers</td>
<td>-0.000395</td>
<td>-0.12</td>
<td>0.9120</td>
<td>1.0000</td>
</tr>
<tr>
<td>sensor-p-detect*avg-replication-on-servers</td>
<td>-0.001714</td>
<td>-0.51</td>
<td>0.6233</td>
<td>1.0000</td>
</tr>
<tr>
<td>shut-down-threshold*avg-replication-on-servers</td>
<td>0.005146</td>
<td>1.54</td>
<td>0.1263</td>
<td>0.9950</td>
</tr>
<tr>
<td>number-of-attackers*mean-vulnerabilities-learned</td>
<td>0.002723</td>
<td>0.82</td>
<td>0.3992</td>
<td>1.0000</td>
</tr>
<tr>
<td>sensor-p-detect*mean-vulnerabilities-learned</td>
<td>-0.001863</td>
<td>-0.56</td>
<td>0.5934</td>
<td>1.0000</td>
</tr>
<tr>
<td>shut-down-threshold*mean-vulnerabilities-learned</td>
<td>0.003242</td>
<td>0.97</td>
<td>0.3179</td>
<td>1.0000</td>
</tr>
<tr>
<td>avg-replication-on-servers*mean-vulnerabilities-learned</td>
<td>-0.001323</td>
<td>-0.40</td>
<td>0.7043</td>
<td>1.0000</td>
</tr>
<tr>
<td>number-of-attackers*hacker-learning-time</td>
<td>0.000426</td>
<td>0.13</td>
<td>0.9049</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Limitations of CIA as KPI
Infrastructure Example Part 1

S (strength: resources of player 1 to defeat player 2)

<table>
<thead>
<tr>
<th></th>
<th>Attack 1</th>
<th>Attack 2</th>
<th>Attack 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>1</td>
<td>3</td>
<td>0.5</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.33</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>Setup 3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

C (cost)

<table>
<thead>
<tr>
<th></th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.23</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.56</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.21</td>
</tr>
</tbody>
</table>

\[ U \text{ (utility)} \]

<table>
<thead>
<tr>
<th></th>
<th>Attack 1</th>
<th>Attack 2</th>
<th>Attack 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.000</td>
<td>-0.043</td>
<td>0.095</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.043</td>
<td>0.000</td>
<td>-0.070</td>
</tr>
<tr>
<td>Setup 3</td>
<td>-0.095</td>
<td>0.070</td>
<td>0.000</td>
</tr>
</tbody>
</table>

- One player loses all utility, another fraction
- Setup 2 vs Attack 1:
  gain - loss
  \[0.23 - (1/3 \times 0.56)\]
# Infrastructure Example Part 2

## U (utility)

<table>
<thead>
<tr>
<th></th>
<th>Attack 1</th>
<th>Attack 2</th>
<th>Attack 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.000</td>
<td>-0.043</td>
<td>0.095</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.043</td>
<td>0.000</td>
<td>-0.070</td>
</tr>
<tr>
<td>Setup 3</td>
<td>-0.095</td>
<td>0.070</td>
<td>0.000</td>
</tr>
</tbody>
</table>

## P (probability)

<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.336</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.456</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.208</td>
</tr>
</tbody>
</table>

## P (probability)

<table>
<thead>
<tr>
<th></th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.333</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.334</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.333</td>
</tr>
</tbody>
</table>

## U (utility)

<table>
<thead>
<tr>
<th></th>
<th>Attack 1</th>
<th>Attack 2</th>
<th>Attack 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.000</td>
<td>-0.073</td>
<td>0.073</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.073</td>
<td>0.000</td>
<td>-0.073</td>
</tr>
<tr>
<td>Setup 3</td>
<td>-0.073</td>
<td>0.073</td>
<td>0.000</td>
</tr>
</tbody>
</table>

## C (cost)

<table>
<thead>
<tr>
<th></th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack 1</td>
<td>0.255</td>
</tr>
<tr>
<td>Attack 2</td>
<td>0.545</td>
</tr>
<tr>
<td>Attack 3</td>
<td>0.200</td>
</tr>
</tbody>
</table>
Resource Acquisition

FLIP IT: The Game of “Stealthy Takeover”.
van Dijk et al. J. Cryptol. (2013)
Lanchester’s Laws

\[
\frac{dA}{dt} = -\beta B \quad \& \quad \frac{dB}{dt} = \alpha A
\]

- X
- o
- X
- o
- X

- X’s expected loss
  - Symmetric Close: 2x
  - Symmetric Ranged: 1x
  - Asymmetric Cyber/Social favoring x: << x
  - Asymmetric Cyber/Social favoring o: <=4x
Information Warfare: Security Through Obscurity

Fortification
Honeypot
Sampling
Adaption

Pavlovic, Proc 2011 ACM New Sec Paradigms Workshop
Surprisal

- Self-information: information of outcome of random event
- Surprisal: $-\log_2 P(x_i)$
- Information: Expected surprisal

![Graph showing surprisal vs. probability]
Information Conveyance
Corpse Party
Chapter 1 Infirmary
Infirmary Flow

- Actual branching factor: 12
- Perceived branching factor: 11
- Exaggerated expectation
  
  [Hilbert, PSYCHOL BULL '12]

- $P(\text{progress} \mid \text{revisit item})$ higher than anticipated
Infirmary Surprisal

• Player unsure of what to do, so assume uniform distribution over new possibilities:

  \[ Q(X) \approx \frac{1}{11}, \quad Q(\text{Repeat}) \approx 0 \quad \Rightarrow \quad \sim 3.5 \text{ bits} \]

• Correct distribution over possibilities, minimizing assumptions: \( P(X) = \frac{1}{12} \)

\[
D_{KL}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}
\]

\( Q(\text{repeat}) \approx 0 \) means

\[
\frac{1}{12} \times \ln\left( \frac{1/12}{0} \right) = \frac{1}{12} \times \ln(\infty) = \infty
\]

Massive surprisal if assume no repeat actions advance game
Measuring Difficulty By Decision Information Rate

3 out of 6 paths: breach

1.5 bits of total information to security breach

1.5 bits / 2 steps = 0.75 bits per step to breach
Exploiting Information Gain

- Measure information gain between current strategy and “optimal”
- Mixed strategy Nash equilibria
  - 1/3 rock, 1/3 paper, 1/3 scissors
- How much information left to learn?
  - 1/4 rock, 1/4 paper, 1/2 scissors
- Info gain to achieve desired Nash equilibrium
Reputation
- Belief about attribute
- Hindsight, capabilities, statistics
- Concern: adverse selection

Trust
- Belief will not exploit
- Foresight, strategy, game theory
- Concern: moral hazard
Trust Heuristic Shortcuts

● Mirroring (homophily)

● Leverage sexual attraction (embedding)
  ● Touch, arm in particular (Guegen, Social Influence, 2007)
  ● Anxiety / fear confused with attraction (Dutton & Aron, J Pers & Soc Psychology, 1974)

● Pretexting (corroboration & homophily)
Mark Selection & Avoidance

● Team sensitivity
  ● Influenced by testosterone
  ● Proxies: index-ring finger length ratio (many studies), width of jaw (Stirrat & Perrett, Psych Science, 2012)

● Sweet eaters = agreeable (Meier et al., J Pers Soc Psychol, 2012)

● Being a "grey man": construction worker, mormon, handicapped, begger, indescribable
"Epidemics" in Social Networks

- Obesity
- Back Pain
- Suicide
- Sex practices
- Politics

Christakis & Fowler, 2007 & 2011

- Information leaks from periphery of organization
Corroboration vs Influence: Some "Epidemics Are From Homophily"

Triangles:
Themes (Eckman & Moses, PNAS 2001)

Aral, Muchnika, Sundararajana 2009
Acceptance and Affirmation

Antitheses (Alliterated)

- Algorithm aversion Dietvorst et al., J Exp Psych 2013

+ Anthropomorphization

- Abeyance of absorption

+ Acceptance acquired after asking assistance

Flynn, Org Behav & Hum Dec Proc, 2003
• **Confirmation Bias**: larger distance between prior belief and new observation \( j \) compared to \( k \) means diminished impact of observation \( j \):

\[
D_{KL}(f_{a,I \cup j}||f_{a,I}) > D_{KL}(f_{a,I \cup k}||f_{a,I}) \\
\land D_{KL}(h_{a,I \cup j}||h_{a,I}) < D_{KL}(h_{a,I \cup k}||h_{a,I})
\]

• **Fundamental Attribution Error**: attribute observation to attributes of actor instead of situation (culturally influenced):

\[
f_{a,I \cup j} = f_{a,I \cup k} = f_{a,I \cup l} \\
\land h_{a,I \cup j} \neq h_{a,I \cup k} = h_{a,I \cup l} \\
\land t_j = t_k \neq t_1 \land s_j \neq s_k = s_l
\]

• **Bias Blind Spot**: don’t know on biases, so can’t correct for them even if they are known (“better than average”)

\[
f_{a,I \cup j} \neq f_{a,I} \land h_{a,I \cup j} = h_{a,I} \land t_j = a
\]

• **Anchoring Bias**: uneven weighting between observations - certain observations stand out more than others:

\[
D_{KL}(f_{a,I \cup j}||f_{a,I}) = D_{KL}(f_{a,I \cup k}||f_{a,I}) \\
\land D_{KL}(h_{a,I \cup j}||h_{a,I}) \neq D_{KL}(h_{a,I \cup k}||h_{a,I})
\]

• **Representativeness Bias**: incorrect propagation of probability:

\[
h_{a,I}(x) \neq \frac{\prod_{i \in I} h_{a,i}(x)}{\int_X \prod_{i \in I} h_{a,i}(y)dy}
\]

• **Projection Bias**: own shortcomings attributed to situation/others rather than self:

\[
D_{KL}(f_{a,I \cup j}||f_{a,I}) > D_{KL}(f_{a,I \cup k}||f_{a,I}) \\
\land D_{KL}(h_{a,I \cup j}||h_{a,I}) < D_{KL}(h_{a,I \cup k}||h_{a,I}) \\
\land t_j = a \land t_k \neq a
\]
Conclusions

• Goals, Economics, and Verbs
• Ambiguity and Information
• AI!

cjhazard@hazardoussoftware.com