The Critical Role of Positive Incentives in Reducing Insider Threat

Andrew P. Moore

Contributors: SEI CERT, SEI Software Solutions Division, SEI Human Resources, SEI Organizational Effectiveness Group, CMU Heinz College/Tepper School of Business
Copyright 2016 Carnegie Mellon University

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[Distribution Statement A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon® and CERT® are registered marks of Carnegie Mellon University.

DM-0004035
Research Objective

Determine influence of workforce management practices on insider threat behaviors

**Negative Incentives**

Workforce management practices that attempt to *force* employees to act in the interests of the organization

**Employee Constraints, Monitoring, Punishment**

**Positive Incentives**

Workforce management practices that attempt to *attract* employees to act in the interests of the organization

**Focus on Employee Strengths, Fair & Respectful Treatment**

Negative incentives *alone* can *exacerbate* the threat they are intended to mitigate*

**Basic Belief:** Organizations need to *explicitly* consider a *mix of positive and negative incentives* to build insider threat programs that are a net positive for employees

**Initial Scope:** Demonstrate value of research in area for insider threat reduction

Three Dimensions of Employee-Organization Alignment

People

Connected @ Work

Organization

Perceived Organizational Support

Job Engagement

Job
Two-Pronged Exploratory Research Approach

1. **Insider Incident Case Study Analysis**
   - How engaged, connected, and supported are insider threat actors?

2. **Organizational Survey**
   - How much does organizational support influence insider cyber misbehavior?

Extension of previous work by focusing on
- Egregious insider threat behaviors
- Organizations actively establishing insider threat programs
Insider Incident Case Study Analysis

How engaged, connected, and supported are insider threat actors?

• **Method:** Rate dimensions on 5-point Likert scales over three time periods
  - For example, for Job Engagement
    
    ![Likert Scale Diagram](image)

• **Challenge:** Assessing insider perceptions through observables (w/o interview)

• **Results:** (3 prominent incidents)
  - Dimensions became increasingly negative over time, with some fluctuation
    - *Organizational Support* most strongly negative in all 3 incidents
    - *Job Engagement* negative in 2 out of 3 incidents
    - *Connectedness at Work* negative in 1 out of 3 incidents

• **Initial Decision:** Focus on perceived organizational support as foundation.
Organizational Survey

How much does organizational support influence insider cyber misbehavior?

**Challenge:** Hard-to-reach population suggests initial exploratory

**Method:** Survey insider threat program managers in an Insider Threat Information Sharing Group

- Independent variable on established 5-point scales
  - *Perceived organizational support* (36 questions)
- Dependent variable on 5-point frequency scale
  - *Cyber misbehavior* from case data (22 questions)

**Response:**

- 25 out of ~90 organizations responded

**Results:** (23 responses used)*

* Analysis used Deming Regression and Multiple Imputation by Chained Equations for missing values.
Emerging Physics of Job Satisfaction, Disgruntled Insider Threat

- Satisfaction levels fairly constant over time. (OPM 2014) (Gallup 2013)

Key:
- A stock (grouping)
- A flow between stocks
- A direct (positive) influence of one variable on another

- ~55% of USG workforce
- ~45% of USG workforce
- ~18% of USG workforce

- Employees Satisfied with Organization
- Employees Dissatisfied with Organization
- Disgruntled Insiders
- Former Disgruntled Insiders
- Former Employees
- Insider Threat Incidents

- Hiring satisfied employees
- Hiring dissatisfied employees
- Employees becoming dissatisfied
- Employees becoming satisfied
- Terminating satisfied employees
- Terminating dissatisfied employees
- Becoming disgruntled
- Reengaging or terminating disgruntled
- Starting to attack
Preliminary Model Simulation Findings

Simulation Controls:

- **Percent disgruntled starting to attack:** 0, 0.2, 5
- **Percent satisfaction improvement:** 0, 50, 100

Sensitivity simulation over the two inputs:

- Number of Insider Incidents After 20 Years
- Precent Satisfaction Improvement

Other model uses: Calculate the cost savings from fewer incidents and less counterproductivity.
Future Research

**Theory Development**

- Experiment-based determination of cause-effect relationship between perceived organizational support and insider threat

**Technology Development**

- Detection of insider alienation by identifying at-risk behaviors and indicative changes in insiders’ network of workplace relationships

**Adoption**

- Determine how particular organization can
  - determine an appropriate mix of positive and negative incentives
  - transition to that from their current state
Vision: Extending the Traditional Security Paradigm

- Fewer unintended consequences
- Satisfaction, performance, retention

Security through Positive Incentives

- Engagement Feedback
- Engagement
- Connectedness
  - Engaged Employees
  - Connected Employees
  - Supported Employees

Traditional Security Approach (Negative Incentives)

- Deterrence Feedback
- Deterrence
  - Deterred Abuse
  - Prevented Abuse
  - Detected Abuse
  - Punished Abuse

- Fewer insider incidents and misbehaviors
- Lower investigative costs, productivity loss

Balanced Deterrence
Contact Information

Presenter / Point of Contact:
Andrew Moore
Lead Insider Threat Researcher
Telephone: +1 412.268.5465
Email: apm@cert.org

Contributors:

SEI CERT:
Samuel J. Perl
Jennifer Cowley
Matthew L. Collins
Tracy M. Cassidy
Nathan VanHoudnos

SEI SSD:
William Novak
David Zubrow

SEI Directors Office:
Palma Buttiles

SEI Human Resources:
Daniel Bauer
Allison Parshall
Jeff Savinda

SEI Organizational Effectiveness Group:
Elizabeth A. Monaco
Jamie L. Moyes

CMU Heinz College and Tepper School of Business:
Professor Denise M. Rousseau

Special thanks to the Open Source Insider Threat (OSIT) Information Sharing Group for their responses to our survey.
Backups
Positive Incentive-Based Principles and Practice Areas

The Critical Role of Positive Incentives in Reducing Insider Threat

October 2016
© 2016 Carnegie Mellon University

(DISTRIBUTION STATEMENT A) This material has been approved for public release and unlimited distribution.
Potential Unintended Consequences of Traditional Insider Threat Management Practices*

1a. Unprotected Whistleblowers Bite Back
1b. Distrusting Whistleblower Goes Public
2a. Aggressive Detection Alienates Employees
2b. Aggressive Prevention Inhibits Effectiveness
2c. Secretive Surveillance Breeds Distrust
2d. Open Surveillance Promotes Evasion/Subversion
3a. Costly InTP Undermined by Perceived Higher Priorities
3b. Ineffective Program Loses Steam
3c. False Positives Erode Support
3d. Too Much Information Erodes Support
3e. Apparent Success Diminishes the Perceived Need
4a. False Accusations Undermine InTP
4b. Abusive Staff Corrupt InTP Function
4c. Anxious Employees Put Up Smokescreen
4d. Opportunistic Managers Monitor Productivity
4e. Overblown Threats Mis-Prioritize Resources
5a. Inconsistent Execution Breeds Unfairness
5b. Investigations Unfairly Affect Employees’ Careers
5c. InTP Detection Allows Accidental Disclosure

Cost Benefits Due to Fewer Counterproductive Behaviors (CWB) and Insider Threat Incidents

Assumptions:

- Satisfied employees engage in one CWB every 2 months
- Dissatisfied employees engage in two CWBs every month (4 times the rate of satisfied)
- Average cost of a CWB is $500
- Average cost of an insider incident is $1M
- Includes data loss, IP loss, investigation costs, etc.
- Cost index calculated by dividing the costs associated with CWBs and incidents by the cost in the baseline run
Workplace Violence/IT Sabotage: Two Sides of the Same Coin?

Presenter: Michael C. Theis
Research Objective and Approach

Objective: Determine if coherent, integrated, and validated indicators for Insider Workplace Violence (WPV) and Insider Cyber Sabotage (ICS) can be identified.

Reason: If there are common indicators organizations may be able to develop socio-technical controls that prevent, detect, and help respond to both threats without identifying which crime will eventually be committed.

Approach: Collect, code, and analyze cases of WPV and compare them to cases of ICS in the CERT Insider Threat Center’s corpus.
WPV and ICS Incident Pathway

- **Hostile Act**
  - Execution of malicious code
  - Active shooter

- **Problematic Organizational Responses**
  - Demotion without changing access
  - Loss/Suspension of Rights and Privileges

- **Concerning Behaviors**
  - Visiting internet underground
  - Verbal threats to cause physical harm
  - Lack of resources to do job well
  - Perceived harassment by coworkers

- **Stressors**
  - Hacker
  - Resolving conflict by physical means

- **Personal Predispositions**

Key:
- WPV
- ICS

CERT, 2006
Hypothesis: Common Path Before Divergence
Aggregation of Stressors

- Personal: 19%
- Work: 39%
- Financial: 9%
- Relationship: 10%
- Mental Health: 7%
- Work Relationship: 16%

© 2016 Carnegie Mellon University
Stressors by ICS and WPV

Stressors for ICS
- Personal: 31%
- Work: 12%
- Financial: 51%
- Relationship: 0%
- Mental Health: 5%

Stressors for WPV
- Personal: 12%
- Work: 21%
- Financial: 8%
- Relationship: 10%
- Mental Health: 36%

Workplace Violence/IT Sabotage: Two Sides of the Same Coin?
26 October 2016
© 2016 Carnegie Mellon University
Distinguishing the WPV and ICS Pathways

Prevalence of Physical Manifestation
- Predispositions
  - Resolving conflicts
  - Perceived harassment by coworkers
- Stressors
  - Verbal threats of physical harm
- Concerning Behaviors
  - Unauthorized weapons at work
- Attack Execution
- Incident Timeline
- Workplace Violence
- Cyber Sabotage

Prevalence of Cyber Manifestation
- Member of hacking community
- Losing control of system evolution
- Visiting hacking websites
- Introduction of malicious code

Average behavior
- Range of behavior

Potential measure to distinguish attack type
Backups
A Cyber-Physical Scale for Assessing Observables*

Examples:
Execution of malicious code to delete critical files
Disabling the code that creates the backup tapes
Not submitting online reports in a timely manner
Other observables that do not reach the level of aggression
Not coming to meetings scheduled
Spreading false rumors
Verbal assault, bullying, shooting coworkers

* Note: combined cyber-physical observables may be broken down into their constituent components for measurement. See the Reality-Virtuality Continuum for a loosely related construct applied to virtual reality technologies. [https://en.wikipedia.org/wiki/Reality%E2%80%93virtuality_continuum](https://en.wikipedia.org/wiki/Reality%E2%80%93virtuality_continuum)
Operational Definitions (from Buss and Parrot)

Aggression – intentional behaviors that can cause significant harm to a victim (person or organization) who wishes to avoid the act. (note: definition excludes desired harm (sadomasochism, going to dentist) and unintentional harm (stepping on foot))

**Direct Aggression** – person-to-person interactions (but not necessarily face-to-face) in which the perpetrator is easily identifiable by the victim (e.g., Active: Shooting, email a threat; Passive: intentionally not write a letter of recommendation and harming victim’s application for new job).

**Indirect Aggression** – circuitous interactions in which the perpetrator may remain unidentified, possibly to avoid accusation, direct confrontation, and/or counterattack by the victim (e.g., Active: (anonymously) spreading false rumors; Passive (rare): (anonymously) not coming to the defense of someone being criticized).

**Active Aggression** – an act of commission by the perpetrator, which involves active engagement in harming the victim (e.g., Direct: shooting; Indirect: (anonymously) spreading harmful rumors)

**Passive Aggression** – an act of omission by the perpetrator, which involves a lack of active responding that causes harm to the victim (e.g., Direct: intentionally not write a letter of recommendation and harming victim’s application for new job; Indirect (rare): (anonymously) not coming to the defense of someone being criticized)

**Physical** - intentional acts involving personal or interpersonal interaction that does not involve cyber

**Cyber** - intentional acts involving interaction with computers, computer networks, or electronic media
### Hasan, Fort Hood – 2009: Concerning Behaviors

<table>
<thead>
<tr>
<th>Major Period</th>
<th>Sub-Period</th>
<th>Direct-Active Cyber Aggression (-3)</th>
<th>Indirect Active Cyber Aggression (-2)</th>
<th>Passive Cyber (Indirect or Direct) (-1)</th>
<th>Center of Scale (0)</th>
<th>Passive Physical (Indirect or Direct) (+1)</th>
<th>Indirect Active Physical Aggression (+2)</th>
<th>Direct Active Physical Aggression (+3)</th>
<th>Sub-Period Concerning Behaviors (non-zero)</th>
<th>Major Period Concerning Behaviors (non-zero)</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘92-97</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>‘98-03</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>‘04-09</td>
<td></td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Sub-Periods of Last Major Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘04-05</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>‘06-07</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>‘08-09</td>
<td></td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Major Period Totals</td>
<td></td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>
### Alexis, WNY – 2013: Concerning Behaviors

<table>
<thead>
<tr>
<th>Major Period</th>
<th>Sub-Period</th>
<th>Direct-Active Cyber Aggression (-3)</th>
<th>Indirect Active Cyber Aggression (-2)</th>
<th>Passive Cyber (Indirect or Direct) (-1)</th>
<th>Center of Scale (0)</th>
<th>Passive Physical (Indirect or Direct) (+1)</th>
<th>Indirect Active Physical Aggression (+2)</th>
<th>Direct Active Physical Aggression (+3)</th>
<th>Sub-Period Concerning Behaviors (non-zero)</th>
<th>Major Period Concerning Behaviors (non-zero)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/04-3/07</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4/07-12/10</td>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>’1/11-9/13</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Sub-Periods of Last Major Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Major Period Totals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>
7-Point Scale Analysis of Results

Key:
- WPV: Red Circle
- ICS: Blue Circle

Graph showing the comparison between Cyber Aggression and Physical Aggression for Hasan, Alexis Wells, Lopez, ICS1, ICS2, ICS3, ICS4, and ICS5.