Deception in Defense of Information Systems

Fred Cohen
Fred Cohen & Associates
The University of New Haven
Sandia National Laboratories
Outline

• Background
  – Fred Cohen
  – Deception in IP

• Cognition & deception
  – Human perception
  – Computer perception
  – System perception
  – Exploiting perceptions

• Models of deception
  – Mathematical models
  – Heuristic models

• Deception Experiments
  – Devices and applications
  – Red teaming exercises
Fred Cohen

• Fred Cohen & Associates
  – 20+ years of information protection leadership
  – Many complex research projects and subjects
  – Consulting for many leading organizations
    • Government and Academic research and advise
    • Corporate consulting at the highest levels
    • Digital forensics, private and LE investigations
    • Strategic intelligence and critical infrastructures
  – Research and Development
    • Evaluating technologies and lines of effort
    • Patented technologies and innovations
Deception in Info Protection

- Time immemorial
  - There has always been deception in nature
  - Humans used deception since their emergence
  - Information protection started 4,000 years ago
  - Early deception in the form of steganography
  - Ongoing deception in warfare
- Deception in information systems
  - Honey pots
  - Lightning rods
  - Deception ToolKit
  - D-Wall
  - Larger-scale deceptions
  - HoneyNet Project
Deception in Nature

• Chest beating as appearance of fitness
  – Apes and humans

• Mock battle between competing members
  – As between antelopes, seals, walruses, people

• Color changes (deception or reflex?)
  – Octopi, chameleon, various fish

• Shape changing
  – Octopi and some squid

• Coloring (deception or evolution?)
  – Orca whales, plants getting bees, etc.
Human Deception

• Languages:
  – Dunnigan and Nofi: concealment, camouflage, false/planted information, ruses, displays, demonstrations, feints, lies, insight
  – Chuck Whitlock introduces 86 specific scams in his book in words used by the scammers
  – Bob Fellows introduces 21 deception classes and 16 reasons people fall prey to them
  – Other authors offer other sets of words

• Deception is common in everyday life
  – Humans live by it - and die by it
Steganography

- Steganos (Hidden) Graphic (Writing)
  - Steganography = Hidden Writing
  - Concealment s.t. Content is not apparent
- >3,000 years old
  - Write on slaves shaved head, grow hair back
  - Recipient knows to shave slave for content
- WWII
  - Microdots, German positional ciphers, etc.
- Modern
  - Hidden in waveforms, packets, files, protocols...
Deception in Warfare

• Sun Tzu

  - "All warfare is based on deception." ["War is an aberration of Dao"]

    - When able to attack, we must seem unable; when using our forces, we must seem inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near. Hold out baits to entice the enemy. Feign disorder, and crush him. If he is secure at all points, be prepared for him. If he is in superior strength, evade him. If your opponent is of choleric temper, seek to irritate him. Pretend to be weak, that he may grow arrogant. If he is taking his ease, give him no rest. If his forces are united, separate them. Attack him where he is unprepared, appear where you are not expected." -- Sun Tzu, The Art of War
More Deception in War

• Many examples of deception in war
  – Pas de Calais instead of Normandy
    • Bodyguard
  – Whaley collection:
    • 67 military deception operations 1914 - 1968
    • Deception is highly correlated to success
      – Do commanders who use deception win?
      – Or do winning commanders also use deception?
  – "Dewar" overviews this in 12 pages
  – "Strategic Military Deception"
    • Describes different views in some detail
Some Recent Works

• Victory and Deceit - Dirty Tricks at War (Dunnigan and Nofi 95)
  – concealment, camouflage, false and planted information, ruses, displays, demonstrations, feints, lies, and insight

• A Note on the Role of Deception in Information Protection (Cohen 98)
  – applying these notions to modern information systems with Deception Toolkit and other methods
Deceptive Defense Examples

• Concealment
  – Concealed services
  – Hard-to-guess passwords
  – Low building profile
  – Path diversity
  – Retaining confidentiality of security status information
  – Spread spectrum

• Camouflage
  – Noise injection
Deceptive Defense Examples

• False and planted information
  – Perception management
  – Rerouting attacks

• Ruses
  – Not widely used

• Displays
  – Weak encryption

• Demonstrations
  – Widely publicized prosecutions
Deceptive Defense Examples

- **Feints**
  - Announced policies of returning fire, etc.

- **Lies**
  - Feeding false information
  - Isolated sub-file-system areas
  - Traps

- **Insight**
  - Model-based situation anticipation and constraint
  - Simulating cyber defense
Recent IT Deception

- Honey Pots (Cheswick94)
- Lightning Rods (Cohen96)
- Deception ToolKit (Cohen98)
- Distributed Analysis and Response (DARE)
- D-Wall (Cohen99)
- Some mathematical properties (Cohen99)
- Ridlr (NPS99)
- The HoneyNet Project (2000)
Deceptive Defense Properties

- Properties that make deception work:
  - Lots of possible ways to attack & some will work.
  - Which way does the opponent use?
  - Deception:
    1. Influence opponent choices to choose in your favor
    2. Detect opponent sooner, react to cut off other avenues
    3. Consume opponent resources

\[ P(A \text{ in } V) \text{ low} \]
\[ P(A \text{ in } D) \text{ high} \]
\[ Ta(D) \cong \infty \]
\[ Ta(R, D) \cong \infty \]
\[ Td(D) \cong 0 \]
\[ Tr(D) \cong 0 \]
Honey Pot Systems

Normal System Normal System Normal System

Normal System Honey Pot Normal System

Normal System Normal System Normal System

P(A in V) low
P(A in D) low
=> Need to attract
P(A in V) low, P(A in D) high
Ta(D) @ ∞, Ta(R,D) @ ∞
Td(D) @ 0, Tr(D) @ 0
Deception Toolkit

- DTK creates deceptive IP services
  - Deceptions intermingled with normal systems
    - Use otherwise unused ports
    - Provide several steps of depth
  - Detects some technical intelligence attempts
  - Misleads intelligence efforts -> more effort needed
  - More depth of deception gains you time
  - Increased visibility of attacks makes react easier
  - Detection -> switch to pure deception systems
  - Works against many insiders as well as outsiders
  - May deter certain types of attacks
Other DTK actions

• Other actions are programmable
  – shut off access to the IP address
  – switch all services to deceptions for the IP address
  – increase response delays to slow attackers
  – run traceback programs and notify via pager
  – coordinate with other DTK response units
  – pull log trails from remote monitoring stations
  – anything you want to program into it
Flexible Distributed Defenses

Attacks on different network elements are locally detected

Defending elements independently assess cause, mechanism, and consequences

Defending elements scale up defenses according to their own perceived needs

Defending elements lessen defenses as the attack wanes

Unaffected systems find out about attacks and implement preventive defenses

Model-Based Anticipation and Constraint

Consequence Analysis

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
Multiple Deceptions 1 Box

Even though it looks like a lot of deception boxes they all operate in a single PC and appear to be different

The observer sees many systems, many of which are actually deceptions

The Reality
Addresses are translated multiple times to allow deception networks to be separated from normal networks, to allow ‘real’ machines to replace low fidelity deceptions, and to allow increased indirection & obscurity.

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
Real vs. Perception

Class B network

D-WALL covers much of the space
Coverage 10.212.*.* mixed

Class B network

10.212.*.*
Another Application

Obscure request

Locate traffic

Request
Using historical Dwall technology, a two step system can translate for 64,000 addresses using 17 computers:

```
<table>
<thead>
<tr>
<th>a.b.<em>.</em></th>
<th>m.n.o.p/0-65535</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.b.0.0-a.b.15.255 translates to x.y.a.b/0,0-x.y.a.b/15,255</td>
<td></td>
</tr>
<tr>
<td>a.b.16.0-a.b.31.255 translates to x.y.a.b/16,0-x.y.a.b/31,255</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>a.b.240.0-a.b.255.255 translates to x.y.a.b/240,0-x.y.a.b/255,255</td>
<td></td>
</tr>
</tbody>
</table>
```

Using a single system, the same translation can be done in software or with custom hardware if very high speed is desired. Implemented, 2001.

Input is examined using 'promiscuous mode, and a translation table is used to associate each internal address to an external address/port number.
• An NPS implementation of a honey pot
  – Secure copy of all data at interface
  – Easy to restore system in standard setup
  – Separated from other systems to prevent unnecessary risk of exploitation

• Regular attacks by many sources
  – Some attacks multi-sourced
  – Observed hand of some users returning
  – A wide range of skill levels
  – Took 3 hours before first attack
The HoneyNet Project

• Goal: To collect data on current attacks
• Method: Honeypots around the Internet
• Activity: Volunteers configure and maintain
  – Attackers attack
  – Defenders observe
  – After a while defenders rebuild systems
  – Defenders do forensic analysis of attacks
  – Early warnings of some major viruses
• The ONLY Deception is that these systems are in the HoneyNet - otherwise normal
Other Deceptive Defenses

- Flexible IP addresses (change every second?)
- Internal deceptions (within the OS)
- High grade deceptions (larger-scale fictions)
- Misleading intelligence (aggressive Opsec)
- Systemic deceptions (embedding deception)
- War modes, feints, etc.
- Strategic deception support
- No end in sight
Break Time
Cognitive Models

• Objective
  – Examine cognitive models of Deception
  – Create a theoretical framework for deception
  – Perform experiments to refute or confirm
  – Use theory to build better deceptions
  – Understand effectiveness in terms of the theory
  – Improve automation and tools for deception
  – Improve counter-deception as an intelligence process
A Human / Human Organization Model of Deception

Intent:
Objectives / Quality
Schedule / Budget

Expectations
Fidelity / Biases
Effort level
Consistency w/
observables

Actions

Capabilities:
Tools
Skills
Methods

Logic & Reason
Pattern Matching

Match
Move
Dissonance

Assessment:
Decisions
Evaluation
Measure Effect

Reason

Observables:
Observable set
Sensor bias
Sensory data

Reflexes
& Conditioned
Response

Deception:= induce
/suppress enemy signals

Logic & Reason
Pattern Matching

O}

Training / Instinct

Capabilities:
Tools
Skills
Methods

Reason

Observables:
Observable set
Sensor bias
Sensory data

Training / Instinct

Deception:= induce
/suppress enemy signals

World
Human Deception Levels

- Low level deceptions:
  - Physiologically derived / well understood
  - e.g., reaction time, edge line detection, recovery time, etc.

- Mid level deceptions:
  - Trained level / less well understood
  - e.g., Pavlov's dog, pattern match limitations, etc.

- High level deceptions:
  - Reasoning and expectation based
  - e.g., short con, big con, grand deceptions

- Group level deceptions:
  - Group dynamics based
  - e.g., social influence, organizational structure, etc.
A Human / Human Organization Model of Deception

High Level

Middle Level

Low Level

World

Deception: = induce / suppress enemy signals

Intent:
Objectives / Quality
Schedule / Budget

Expectations
Fidelity / Biases
Effort level
Consistency w/ observables

Assessment:
Decisions
Evaluation
Measure Effect

Reason
Capabilities:
Tools
Skills
Methods

Logic & Reason
Pattern Matching

Match
Move
Dissonance

Observables:
Observable set
Sensor bias
Sensory data

Reflexes & Conditioned Response
Actions

Training / Instinct

Capabilities:
Tools
Skills
Methods

Reason

Logic & Reason
Pattern Matching

Decisions
Evaluation
Measure Effect

Observables:
Observable set
Sensor bias
Sensory data

Reflexes & Conditioned Response
Actions

Training / Instinct
Deception Techniques by Level

• Low Level
  – Identify physiological limits / factors
  – Identify methods to exploit them
    • Induced signals at observation points
    • Suppressed signals at observation points
  – Create capabilities to exploit them
  – Exploit them as needed for effect

• Example:
  – Optical Dazzlements for blinding (lasers)
Sensory limitations

- Limited sensitivity ranges
  - a few dBs < hearing < 150 dBs (too loud)
  - Touch sensitivity limited by pressure range

- Limited frequency ranges
  - Ears: 20Hz<hearing<20,000 Hz
  - Eyes: infrared < sight < ultraviolet

- Limited differentiation
  - You can only taste/smell select molecules

- Saturation and recovery phenomena
  - Flashbulbs temporarily blind you
Perception Limitations

• 5 Senses available (+ pheromones and allergies?):
  – Habituation: You tend to ignore what you sense repeatedly without change
    • The motor noise from the projector
  – Inhibition: High levels of sensation at one receptor tends to reduce sensitivity to nearby receptors
    • Loud noises dominate soft noises
  – Hernandez Poen effect: One sense tends to dominate another at high sensation level
    • Close your eyes and you hear better
Focus of attention

- Sensory input focus is sequenced
  - Priority interrupt system
    - Edge line detection / Startle responses
  - Perceptual sequencing
    - Trained sequences (pilots)
    - Common early learning sequences (how to listen)

- Limits on simultaneous attention
  - Sensory overloads
  - Dazzlements
  - Distractions
  - Masking and aliasing
A simple human deception

- Disappearing elephant
  - Crowd distracted by startle reflex
  - Concealment from sensors by a curtain
- Startle produces reflexive response
  - Crowd turns toward crash noise for 1 sec.
  - Black 'wrinkling' curtain rises in front of elephant. Rear wrinkling curtain conceals movement by removing edge line detection

```
Crashing noise (simulated threat)  Heads turn to noise (ref exive response)
Sense  Induced Reflex  Response
Sense  Move curtain (conceal movement by turned heads and no edge lines)  Feature
Sense  Elephant Hidden (concealed by curtain)
```
Sensory / Other Triggers

• Frequencies synchronized to colon
  – Cause crowds to have urgent bathroom needs
• Frequencies to 'spook' people
  – Cause people to believe there are ghosts
• Frequencies for discomfort
  – Cause pain, nausea, dizziness, etc.
• Frequencies for distraction
  – Other frequencies cause other affects
• Other environmental factors affect sensors
  – Oxygen, blood sugar, temperature, etc.
Low-Level Examples

- Authentication of target characteristics
  - Use color sensitivity for authentication
    - Specific color blindness will see different items
- Dazzlement
  - Use colors in patterns to disrupt vision
    - Cause edge line confusion for observer
- Concealed movements
  - Use smoke to cover movement from vision
    - Prevent visual information from reaching observer
- Heat the room the person is in
  - Use temperature to slow reflexes and exhaust
Low level action limitations

- **Limits on physiology**
  - Speed, force, acceleration, compliance, delay time, throughput, simultaneity, reach, accuracy

- **Limits on focus and sustainability**
  - Muscle fatigue and recovery
  - Focus of attention, fatigue, and recovery
  - Exhaustion, strain, sprain, breaks, collapse
  - Low blood sugar, low oxygen

- **Limits of accuracy**
  - Affected by use, training, history, practice, skill levels, 'natural ability', coordination
Deception Techniques by Level

- Middle Level
  - Train/study the target / test responses
    - e.g., repeated exercise of response methods
  - Wait/study till responses are predictable
    - e.g., every Tuesday 10AM is a fire alarm practice
  - Create false/known rejection/acceptance
    - False rejection exploited by real incident
    - Known acceptance used to induce desired behaviors
  - Exploit conditioned response
    - e.g., use training against them
    - e.g., use training to identify / authenticate
A Human / Human Organization Model of Deception

Intent:
Objectives / Quality
Schedule / Budget

Expectations
Fidelity / Biases
Effort level
Consistency w/
observables

Match
Move
Dissonance

Assessment:
Decisions
Evaluation
Measure Effect

Reason
Capabilities:
Tools
Skills
Methods

Logic &
Reason
Pattern
Matching

Capabilities:
Tools
Skills
Methods

Assessment:
Decisions
Evaluation
Measure Effect

Reason

Logic &
Reason
Pattern
Matching

Capabilities:
Tools
Skills
Methods

Assessment:
Decisions
Evaluation
Measure Effect

Reason

Logic &
Reason
Pattern
Matching

Capabilities:
Tools
Skills
Methods

Training / Instinct

Observables:
Observable set
Sensor bias
Sensory data

Ref exes
& Conditioned
Response

Actions

World

Deception:= induce
/suppress enemy signals

Deception:= induce
/suppress enemy signals

Training / Instinct

Observables:
Observable set
Sensor bias
Sensory data

Ref exes
& Conditioned
Response

Actions

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
Sensory Habituation

- Sensors start ignoring repeated inputs
  - Mid level functions tire of forcing added attention
- Sensors or response requirement overwhelmed
  - Mid level functions move toward pattern matching
- Focus of attention forced by training
  - Mid level functions learn how to sequence attention
- Sequencing of sensors is predisposed
  - Tendency to sequence signals by known sequences
- Prioritization scheme is predisposed
  - Tendency to prioritize vision over sound, etc.
- Slow change (the boiling frog)
Preparing predisposition

- Patterns tend toward predisposed matches
  - Once a theory is predisposed it is hard to see others
- Experiential predisposition in face of ambiguity
  - Tendency toward previous experienced patterns
- Organization of information by previous bias
  - Tendency to organize new things as old things
- Memory / association limits force structuring
  - 7±2 memory chunking phenomena
- Contextual association
  - Provided a plausible theory, people tend to believe it
Pattern matching limits

- Like as like
  - Tendency to associate similar things together
    - Mode matching (Satire's modes)
      - Sound good to you? It sure resonates with my ideas.
      - He'll torpedo me (to a seaman).
    - Tendency for dissonance when like as not like
      - Mode mismatch
        - See what I mean? I hear you.
        - He'll torpedo me (to an air force pilot).
  - Oversimplification tendency
    - Occam's razor?
      - The simplest theory that fits is always preferred
    - Black and white vs. shades of gray
Pattern matching limits 3

- Misperception of random events
  - Clustering illusion
    - Clusters from randomness (pictures in clouds)
  - Law of small numbers
    - Small numbers of examples are over-applied
  - Misperception of random dispersions
    - Shooting streaks
  - Creation and adoption of causal theories
    - Adopt theories to explain, retain in face of refutation
  - Regression fallacies
    - After a big week, you seem in a slump when returning to normal
Focus of attention limits

– Attention span
  • People tend to have limited attention spans
  • Attention moves from sensor to sensor
  • Individual sensors move to observe / attend

– Tendency to listen to novelty
  • New inputs are more attended to
  • Larger differences drive more attention
  • Novel features attract attention (shapes, etc.)
  • Slow changes not noticed

– Tendency to follow through
  • Tendency to obsess / compulse / doze off
Verbal Abuse Self-Defense

**Self-defense process:**
1) Detect attack
2) Characterize it
3) React appropriately
4) Follow through

<table>
<thead>
<tr>
<th>Sensory Modes: (see, smell, hear, taste, feel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>oD - match modes -&gt; like and agree</td>
</tr>
<tr>
<td>oN - no modes -&gt; neutral</td>
</tr>
<tr>
<td>oM - mismatch modes -&gt; dislike, clash, slow resolution</td>
</tr>
</tbody>
</table>

**Satire's Modes:** (*mismatch between beliefs and expressed beliefs*)
- Blaming (all, none, every...) match -> fight
- Placating (you are right boss) match -> unproductive delay
- Computing (generalities/abstractions) match -> slow productive delay
- DistRACTive (flip from one to the other) match -> hester hester
- Leveling (simple truth as they see it) match -> honesty - not always good

**Assumptions & baiting & harsh emphasis -> attack**
- Ignore bait (even you could do that)
- Find presuppositions (you are incompetent)
- Transmit 'it won't work', 'I won't play' (ignore bait)
- Known how to follow through (ask 'when' lever mode)

<table>
<thead>
<tr>
<th>Everybody knows... and we understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&gt; I'm sure they do understand and I appreciate it.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>If you cared about X, you wouldn't Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&gt; When did you come to think I didn't care?</td>
</tr>
</tbody>
</table>

**Behavorial inconsistencies have causes**
Anything you feed will grow
Anything you suppress will fester or die

**Miller's Law:**
- Assume they are telling the truth
- Figure out what they are telling the truth about

**Techniques to use:**
- Presupposition to avoid apposition to generate assumptions
- Illusion of choice when there is none
- 3-part message constant use of blaming, placating, or distraction

**When you do X, I feel Y because Z**

<table>
<thead>
<tr>
<th>Everybody knows... and we understand</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&gt; I'm sure they do understand and I appreciate it.</td>
</tr>
</tbody>
</table>

**If it is general, agree in general**
- ... anyone who would X should/is Y
- ... some people... any fool could ... .or you're not the only X that Y

<table>
<thead>
<tr>
<th>If you cared about X, you wouldn't Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>=&gt; When did you come to think I didn't care?</td>
</tr>
</tbody>
</table>

<p>| =&gt; I agree                               |</p>
<table>
<thead>
<tr>
<th>Cause</th>
<th>Emotion</th>
<th>Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frustration</td>
<td>Hostility</td>
<td>Aggression/apathy</td>
</tr>
<tr>
<td>Threat</td>
<td>Fear</td>
<td>Fight/flight</td>
</tr>
<tr>
<td>Conflict</td>
<td>Anxiety</td>
<td>Inefficiency</td>
</tr>
<tr>
<td>Violation of values</td>
<td>Guilt</td>
<td>Arbitrary Rejection</td>
</tr>
<tr>
<td>Loss</td>
<td>Sorrow</td>
<td>Crying</td>
</tr>
<tr>
<td>Failure</td>
<td>Self-pity</td>
<td>Overindulgence</td>
</tr>
</tbody>
</table>
Preconceptions

- Preconceptions induce selective attention
  - Predisposition creates expectations to
    - Hear what you expect to hear
    - See what you expect to see
  - Interaction of expectations with sensations
    - Literally hear / see / feel what you expect to
      - Await expected sensations
      - Tend to treat similar sensations as those expected
      - Tend to ignore or discount contrary sensations
      - Tend to generate expected sensations on your own
Deceptive authentication

• Conditioned response to screen content
  – Individually train each user to hit a particular key sequence when encountering a particular screen sequence
    • When you see a flash of green in the upper left corner followed by a flash of red 0.2 seconds later in the lower right corner, press the control key with your left little finger knuckle within .04 seconds.
  – It is covert because it is almost invisible
  – Even if you know of it, it is hard to do
    • Until you have been trained
  – Many individual stimulus / response pairs
Deception Techniques by Level

• High Level (What will they accept?)
  – Negotiate a sequence of contracts
    • Create expectation to change trust
  – Fulfill each to increase level of trust
    • Meet/exceed expectations => move expectation
  – When trust is high enough, exploit
    • Expectation leads to positional advantage
  – Get away / extortion
    • Cut off relationship while holding advantage
    • Move expectations from trust to fear
A Human / Human Organization Model of Deception

- Intent: Objectives / Quality Schedule / Budget

- Expectations
  - Fidelity / Biases
  - Effort level
  - Consistency w/ observables

- Actions
  - Observables: Observable set
    - Sensor bias
    - Sensory data

- Logic & Reason
  - Pattern Matching

- Observables:
  - Observable set
  - Sensor bias
  - Sensory data

Deception:= induce / suppress enemy signals
Reasoning limits

- Misrepresentation of incomplete data
  - Excessive impact of confirmations
    - Small number of confirmations taken as proof
    - Refutations ignored or explained away
  - Tendency to seek confirmations
    - Pattern matching configured for target detection
    - Other targets ignored / non detection ignored
  - Hidden or absent data problems
    - Non repeatable experiments: you don't know what would have happened in the path not taken.
- Self fulfilling prophecies
  - Market crashes
Reasoning limits 2

- Biased evaluation of ambiguity and inconsistency
  - Ambiguous data interpretation in context
    - Tendency to find things like what you look for
  - Unambiguous data shaded
    - Tendency to explain away falsifications
  - Multiple endpoints problem
    - Ambiguous data associated with expected outcomes
  - Confirmations and non-confirmations
  - Focused and unfocused expectations
• Biased evaluation (continued)
  – Outcome asymmetries
    • Hedonic: Overemphasis on more striking things
      – Seems more informative if more unusual or stranger
    • Pattern: Overemphasis of specific patterns
      – Remember 1:11 more than 3:46
    • Definitional: Loose definitions / interpretations
      – You won't get better till you hit 'rock bottom'
      – If a tree falls in a forest and nobody is there...
    • Base rate: You only measure survivor views
      – Say "80% of Cancer survivors 'thought' healthy thoughts"
      – Since you cannot ask those who died, you do not know whether healthy thoughts were correlated to survival. 90% of those who died may have thought healthy thoughts.
• Motivational determinants of belief
  – Empirical support for the wish to believe
    • Interpreting the same information in different ways
      – After the Nixon / Kennedy debates, supporters for both sides said that they believed that they won
  – Mechanisms of self-serving beliefs
    • Believers ask "Can I believe?"
    • Non-believers ask "Must I believe?"
  – Optimistic self-assessment
    • Most people believe they are above average
      – In beauty and in mental capacity
Reasoning limits 5

• Biasing of second hand information
  – Sharpening and leveling
    • People emphasize (sharpen) focal points
    • People de-emphasize (level) side points
    • Focal vs. Side depends on the interpreter
  – Corruption with transitivity (game of telephone)
  – Telling a good story (enhancing reader interest)
  – Distortion for informativeness (exaggeration)
  – Distortion for entertainment (humor/interest)
  – Distortion for self interest (greed)
  – Distortion for plausibility (urban legends)
Emotion effects cognition

– Affects:
  • Likes, dislikes, fear, happiness, etc.
    – Positive affect improves sensory detection and recall

– Values:
  • Fairness, right and wrong, etc. impact interest
    – Tendency to be more interested in 'good' things

– Needs:
  • Lack of air, water, food, drive sensor focus
    – Tendency to see food in randomness when hungry

– Interests:
  • More interest leads to better learning
Human perception limits:
High Level: Rand 1999 / MOP 113

- pre-existing notions given excessive weight
- desensitization degrades vigilance
- generalizations or exceptions based on limited data
- failure to fully examine the situation limits comprehension
- limited time and processing power limit comprehension
- failure to adequately corroborate
- over-valuing data based on rarity
- experience with source may color data inappropriately
- focussing on a single explanation when others are available
- failure to consider alternative courses of action
- failure to adequately evaluate options
- failure to reconsider previously discarded possibilities
- ambivalence by the victim to the deception
- confounding effect of inconsistent data
Deception Techniques by Level

● Group Level (Social Engineering)
  – Understand group dynamics and situation
    • Organizational structure and characteristics
  – Understand who to move where in the group
    • Membership power and influence roles / relations
  – Create social pressures for movement
    • Influence key decision makers
    • Busy or deal with key impediments
    • Form group momentum shifts
  – Support movement as desired
    • Create feedback to support desired movements
Organizational Structures

- **Hierarchy**
  - Like the military, a tree structure
    - Move branches by moving their roots

- **Star**
  - Everyone reports to one person
    - Move central focus to move the rest

- **Matrix**
  - Typically product managers vs. line managers
    - Move rows / columns to get intersections

- **Network**
  - Typical of loose-knit organizations
    - Move confluence by moving influential people
Social Pressures

- Reinforcement
  - Punishment and rewards for behavior
    - Reward for conforming in raises, stature, etc.
    - Punishment for nonconformance by shunning

- Imitation
  - Performance improves with similarity to group
    - Behaving like group enhances performance
    - Perception like group helps behave like group
  - Punishment for not acting like group
    - Learning what to do (imitation)
    - Learning what not to do (inhibition)

- Authority and stature induce conformance
- Education and ethics reduce conformance
Introduce as an expert and you will be believed as one
Unless you're damned sure, say I reckon - Media may lend credibility

Credibility

Learning comes from hearing & understanding

Change comes from learning & acceptance

Acceptance comes from comfort with the message & emotion

Audience motives & values
Audience information & language

Audience perception & role
Audience attitudes & emotions

Media choice

situation setting & rewards

Media choice

Message content & appeal

Present both sides favored viewpoint last start / end remembered end remembered best state conclusions repetition helps arouse need then fulfill threats are rejected desirable message first ask for more, get more stress similarities tie hard issues to easy don't create defensive don't belittle other views friendly/sympathetic ask advice appeal to self-worth, fairness, excellence

Letters are good when establishing justification or to get a letter back or when interruption is dangerous Face to face is better when presence brings regard/respect - when visual indicators will help - when more or less may be desired

References

Tend to resolve ambiguity quickly
Social forces account for audience facts, methods, goals, and values
Power issues

People like balance
Ambiguity upsets tendency to resolve ambiguity quickly
Social forces account for audience facts, methods, goals, and values
Power issues

Make the audience feel worthwhile reinforce opinions
Negotiation tactics*

• Timing
  – Patience, deadline, speed, fait accompli, surprise, status quo, stretchout

• Inspection
  – Open, limited, confession, qualified, third party, no admittance

• Association
  – Alliances, associates, disassociates, United Nations, Bribery

• Authority
  – Limited, approval, escalation, missing man, arbitration

• Amount
  – Fail and reasonable, Bullwarism, nibbling, budget bogey, blackmail, escalation, intersection, non-negotiable, Chinese auction

• Brotherhood
  – Equal, bigger, smaller, long-lost, brinkmanship

• Detour
  – Decoy, denial, withdrawal, good and bad guys, false statistics and errors, scrambled eggs, low balling, scoundrel
Timing

- Patience
  - Wait as long as it takes to win

- Deadline
  - Find the deadlines & exploit opponent deadlines

- Speed
  - Quick agreements on small points combine to generate pattern of agreement

- Fait accompli
  - Alter balance of power by already being in place

- Surprise
  - New conditions added during process

- Status quo
  - That's how it has always been

- Stretchout
  - Deliberate delays
Inspection

- Open
  - Unlimited
- Limited
  - Limit in time, location, quantity
- Confession
  - Full disclosure of all circumstances
- Qualified
  - Questions answered but faults not offered
- Third party
  - Independent third party inspectors
- No admittance
  - No inspections
Association

• Alliances
  – Strong partners

• Associates
  – Friends

• Disassociates
  – Mutual non-friends

• United Nations
  – Broad-based association of industry members

• Bribery
  – Payoff and collusion
Authority

- **Limited**
  - I can't make the final decision

- **Approval**
  - I need an approval for that

- **Escalation**
  - I need to show that to my boss

- **Missing man**
  - The person we need to settle that is not here right now

- **Arbitration**
  - Third party - neutral or biased
- Fair and reasonable
  - Exploit morality
- Bulwarism
  - Take it or leave it
- Nibbling
  - Small concessions after issues are thought settled
- Budget bogey
  - I only have so much
- Blackmail
  - No other source
- Escalation
  - Demand once agreed
- Intersection
  - Tie issues together
- Non-negotiable
  - Cannot alter this
- Chinese auction
  - Opponents played off
Brotherhood

• Equal
  – Equal status

• Bigger
  – Benevolence based on higher status

• Smaller
  – Charity based on lower status

• Long-lost
  – Search for relationship and status

• Brinksmanship
  – Intersecting destiny based on shared risk
Detour

- Decoy
  - Attract or snare
- Denial
  - Retraction
  - Walk away
- Withdrawal
- Good and bad guys
  - Good cop bad cop
- Scoundrel
  - Never-ending talk

- False statistics/errors
  - Deceptive statistics
- Scrambled eggs
  - Causing deliberate confusion in figures and statistics
- Low balling
  - Low base expensive add-ons
Power is used to produce Influence

Physical Resource → Overt force
Position information → exchange (one-time OK, repeated⇒expectation)
Expertise personal (charisma) → rules & procedures (perceived right, enforcement)
Emotion (overt) → persuasion (weight depends on belief in source)

Physical Resource → Covert ecology (control environment)
Expertise personal (charisma) → magnetism (highly relative)
Emotion (overt) → Covert ecology (control environment)
Position information right to access right to organize → Covert ecology (control environment)

Bridging threat of force

Adjusting to influence:
- compliance (no choice⇒resentment)
- identification (like idea/person⇒keep recharging)
- internalization (adopt as own⇒ownership changes)

Noise impairs performance
Variety relieves monotony
Seating effects interaction
Layout effects communication
Segregation inhibits communication
Danger increases tension
Smaller groups easier to participate
Attainable challenge⇒commitment
Worthwhile challenge⇒commitment
Interaction increases sentiments

Reason, friendship, coalition, bargaining, assertiveness, sanctions, higher authority
Greater position or resource power ⇒ more strategies used
Reason on bosses, other methods on subordinates is common
More power distance ⇒ fall back on assertiveness is common
Reason is used most when expectation of success is high

Power is relative to the thing being influenced
Balance of power is achieved in most influence
Power is relative to the domain of influence

Credibility in context
Multi-thread stronger

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
Influence and Compliance

- **Reciprocation**
  - Costs more => worth more
- **Authority**
  - Experts know more
- **Contrast**
  - Substantial differences tend to be exaggerated

- **Reciprocation**
  - People tend to reciprocate any gifts
- **Reject and retreat**
  - Ask for something then lower request
- **Commitments**
  - Are honored
- **Consistency**
  - Highly valued
More Compliance

- **Automaticity**
  - Desire not to think
  - Strong desire not to rethink
  - Default decision process
  - Because
  - Enhanced by rush, stress, ..

- **Commitment**
  - Small commitments lead to big ones
  - Active commitments better than passive

- **Commitment**
  - Public image leads to self-image
  - Increased compliance with investment
  - Consistency causes decisions

- **Social proof**
  - Interpret as others do
  - Replaces hard proof in uncertainty
More Cialdini

• Authority
  – Cultural duty to authority
  – Appearance => authority

• Liking
  – Say yes to who you like
  – Physical attraction +
  – Similarity +
  – Compliments +
  – More contact +
  – Groups together bond
  – Groups competing hate
  – Associate with things that enhance self-image

• Scarcity
  – Scarcity implies value
  – Loss > Gain
  – Want restricted stuff
  – Have it our way
  – Exclusive info more valued
  – Drop from abundance => more valued
Example: Influence

- Find key decision makers on sides of issue
  - Zealots: push the desired view
  - Supporters: agree but do not push it
  - Neutrals: have no opinion or don't care
  - Detractors: push against the desired view
- Decide how many to move where to win
  - Move a few detractors to neutral to win?
- Create influences to achieve movements
  - Busy some detractors on higher priorities
  - Persuade some detractors by reason
  - Eliminate some detractors by positional power
• Create a set of 'independent' sources
  – Internet personae with good mix of
    • Organizational affiliations
    • Professional qualifications
    • Backgrounds and experience
    • Positions in organizations

• Affiliate them with a group over time
  – Infiltrate the group in sequence w/bootstraps

• Influence group thinking by combined view
  – Create group think and social pressures
Example: Roving Gang

• Group of professional criminals
  – People with different backgrounds/skills

• Infiltration tactics for access
  – Get a foothold, recommend, hire, build

• Rip off all you can
  – Start taking in large numbers
  – Augment your group

• Move on in timely fashion
  – Find next target early on
  – Begin the move before detection
Deception Techniques
by Level*

- Mixed Level
  - Battle of the Bulge
  - Normandy invasion Bodyguard of lies
  - Egypt Invasion of the Sinai - Suez War
- Target is enemy leadership
- Generally starts with high level plan
- High level needs drive mid level plan
- Mid level needs drive low level plan
A Human / Human Organization Model of Deception

Expectations
Fidelity / Biases
Effort level
Consistency w/ observables

Actions

Intent:
Objectives / Quality
Schedule / Budget

Capabilities:
Tools
Skills
Methods

Logic & Reason
Pattern Matching

Reason
Assessment:
Decisions
Evaluation
Measure Effect

Control:
Match
Move
Dissonance

Observables:
Observable set
Sensor bias
Sensory data

Reflexes & Conditioned Response

Logic & Reason
Pattern Matching

World

Deception:= induce / suppress enemy signals

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
### Command stress and deception

<table>
<thead>
<tr>
<th>Command Baseline</th>
<th>Stressors</th>
<th>Deception</th>
</tr>
</thead>
</table>
| Command Arrangement  
  Line of control  
  Decision process | High stakes  
  Uncertainty  
  Time pressure | Active  
  A-type (ambiguity)  
  repeated feints  
  camoflage  
  false information  
  M-type (misdirection)  
  feints  
  false information | Passive  
  Opsec  
  Camoflage  
  Concealment |
| Expectations  
  Perceptual biases  
  Cognitive biases | | |
## Battle of the Bulge

<table>
<thead>
<tr>
<th><strong>Allied Command Baseline</strong></th>
<th><strong>Stressors</strong></th>
<th><strong>Deception</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Offensive expectations</td>
<td>High stakes</td>
<td>M-type (misdirection)</td>
</tr>
<tr>
<td>(cognitive bias)</td>
<td>For Germans</td>
<td>Encourage allied view</td>
</tr>
<tr>
<td>View of German command</td>
<td>Not for Allies</td>
<td>of German defensive</td>
</tr>
<tr>
<td>situation (perceptual bias)</td>
<td>Maneuver warfare</td>
<td>intent (false info)</td>
</tr>
<tr>
<td>View of German intentions</td>
<td>Uncertainty</td>
<td>Opsec</td>
</tr>
<tr>
<td>(cognitive bias)</td>
<td>Time pressure</td>
<td></td>
</tr>
<tr>
<td>View of German capabilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(perceptual bias)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Normandy invasion

<table>
<thead>
<tr>
<th>German Command Baseline</th>
<th>Stressors</th>
<th>Deception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Command Arrangement</td>
<td>Time pressures: Air attacks Maneuver warfare</td>
<td>A-type (ambiguity) Expected landing at Pas de la Calais (false information, predisposition)</td>
</tr>
<tr>
<td></td>
<td>Consequential decisions: Russian Offensive</td>
<td>M-type (misdirection) German turned spies exploited for Patton invasion (false information, pre-disposition)</td>
</tr>
<tr>
<td>Line of control</td>
<td></td>
<td>Passive Opsec</td>
</tr>
<tr>
<td>(Hitler too dominant)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptual biases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive biases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
# Israel / Egyptian Suez War Surprise

<table>
<thead>
<tr>
<th>Israeli Command Baseline</th>
<th>Stressors</th>
<th>Deception</th>
</tr>
</thead>
<tbody>
<tr>
<td>Command Arrangement</td>
<td>Consequential decisions political cost of mobilizing before elections</td>
<td>M-type (misdirection) - encourage perception of Arab inferiority</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- reflexive conditioning by repeated feints as exercises in Suez</td>
</tr>
<tr>
<td>Expectations</td>
<td></td>
<td>- Syrian / Lebonese air battle</td>
</tr>
<tr>
<td>Air and ground combat superiority</td>
<td></td>
<td>- Australia terrorism?</td>
</tr>
<tr>
<td>Attack warning</td>
<td></td>
<td>Opsec was very well done</td>
</tr>
<tr>
<td>Arab disunity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited # of operations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>at one time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What we don't know about human deception ($\uparrow$)

- Detailed metrics
- Thresholds and limits for moves
- All relevant fields?
  - Brain researchers
  - Computer HW/SW
  - Legal
  - Psychology
  - Experienced teams

- Target information
  - Intel capabilities, culture, thought processes, intent, attack capabilities

- Process
  - Idea generation and selection
  - Team members who develop content
  - Logistics tail
Lunch
Applying Cognitive Deception to Information Protection
Some applications

- A combination of factors that differentiate one person from another leads to authentication
  - Physiological differences (e.g., response times)
  - Trained response differences (e.g., flash and hit)
  - Cognitive perceptual differences (e.g., color blind)
  - Contextual perceptual differences (e.g., situational)
  - Cultural biases (e.g., meaning of select phrases)

- Different schemes for different contexts
  - Initial / Periodic / Random / On Demand
  - Different surety levels
More applications

- Cognitive differences can be used to affect access to / or utility of content
  - Specific color blindness allows visualization
  - Create predisposition to communicate specific content in "cognitive steganography"
  - Create environmental factors to distract
  - Create inconsistencies to get detectors to portray real systems as deceptions
  - Use set points of information availability to induce cognitive overload or underload
  - Train for different set points
Low-level Examples

- Run at rates too high for unskilled users
- Create dazzlements most users can't watch
- Habituate users to hard-to-learn processes
- Create distracting content overcome only by training in what to ignore
- Create dimmed or directional screen effects to reduce shoulder surfing
- Use predisposed observation pattern to derive meaningful content
Low-level Examples

- Hide content using color-blind selectivity
- Hide content in tonal variations
- Hide content with steganographic means
- Create sensory overloads for those who don't know what to ignore
- Create 'too many things' at once for the untrained observer to pick out the proper sequences of observables
Mid-level Examples

• Predisposition
  – If it looks like a duck...
  – It's just DNS traffic
  – Just a search engine

• Trained behavior
  – Duck duck duck Goose
  – DTK's deception port

• Slow change
  – The boiling frog and IDS
Sample problem 1

• Mission: collect data on Internet attacks
  – Turn it into an intelligence capability for systems administrators

• Targets: sites/groups publishing attacks
  – Web sites, IRC chat groups, etc.

• Deception goal: concealment
  – the subjects are unaware of the reason their information is sought - even if they try hard

• Method: plausible explanation
  – They will be led to believe we are something else that just happened across them.
A simple deception:

- Time picture of the deception - single thread case:
  - Threat is over-the-wire only: low-medium resource (e.g., attack sites)
  - We collect data from their locations
    - Try to do it covertly?
      - If they detect it, they will be suspicious and a very good deception will be needed to counter their perception of it as very hostile
    - Try to be overt but not excessive
      - Act like something they expect to come past (e.g., a web spider or some such thing).
      - If they detect it, they may get curious
    - They detect the collection
      - They ignore it, we get the data, they don't know why
      - They attack back / They trace it back
        - Make it innocuous to a level of credibility that they won't question.
        - Make it as consistent as possible with the previous pulls
Diagram of a simple decision process we and they might use
Cognitive Structure

Internal Images
Self (system) image - World image - GOAL - they see a web crawler or two
Knowledge, including data, theories, models of people, physics, etc.

Define problem / observe:
- Policy
- Situation

Define problem solving status: Hypothesize
- Policy met?
- Have tools?

Determine solution options
- investigate?
- attack back?
- refuse service?
- ignore?

Initiate actions / response
- task resources

Associate:
- intelligence pull?
- web crawler?

Perceive form:
- looks high volume
- looks like web crawler

Perceive Feature:
- lots of requests
- unusual requests

Sense:
- requests arrive

Stimulus (presentation)

Detect

React

World Feedback

Intuition

Conditioned behavior

Implement form:
- detail moves OR
- serve / log / warn

Implement feature:
- sequence moves

Drive:
- send responses
- implement moves

Direct:
- order actions

Response (action at a distance)
operator actions
Low-level processes

Perceive form:
- Looks high volume
- Looks like a web crawler
- Automated responses may be used
- Most web sites want to be crawled
- Some web sites view as an intrusion

Perceive Feature:
- Normal logging of web server use
- Automated correlation or presentation may also be done
- Feature extraction may be done:
  - Large volume request sites
  - URLs intended to catch crawlers
  - Cookies and URL-based tracking
  - Referrer fields

Sense:
- Requests arrive along with many other requests of a similar sort
- Web server (as an example) allows normal web requests at a finite rate
- Reflexes send back results unless rate exceeds or otherwise directs

Conditioned behavior
Can we condition a desired set of responses?
- Create expectations of benefits
- Create hassles for hostility
- Create environmental expectation
- Stimulate to find normal response
- Detect response thresholds and can we stay below normal correlation thresholds?

Implementation form:
- Web crawler response option
- Intrusion response option
- Detailed move as directed
- Warn or correlate

Implementation feature:
- Sequence moves for responses
- Prepare warning messages
- Prepare correlation messages
- Alter reflexive responses

实施形式:
- Web crawler response option
- Intrusion response option
- Detailed move as directed
- Warn or correlate

Drive:
- Fulfill web requests
- Implement other responses
- Alter reflexive responses

Response (action at a distance)
Operator actions
Our (desired) prequel

- Do we know enough? (each step)
- Do they have obvious tracking mechanisms?
  - Cookies / special URLs / referrer requirements?
- Their thresholds of detection (experimental)
  - Try to push them to response levels in a major pull from cover sites - what are the responses?
  - Use flexible IPs for cover sites - DHCP via ISPs
  - Real cover sites are expensive - but cutouts?
  - How often can we do it? Are they even looking?
- Their decision process (observe/effect)
  - join them undercover / create crawler benefits / do 'media interviews' / elicitation techniques / technical attacks

- Resource limitations (theirs / ours)
Postquil

– We need to protect the back end from feedback

• When we send this information to our systems administrators, they need to use it but not share it
  – If they share it, it could be traced back to the collector
  – Rewriting content may prevent covert return channels

• Reflexive controls/Trojans by pushing in deceptions

• Obvious disassociation
  – Don't provide detailed collection information to customers
  – Try to obfuscate collection net from customer net
  – Careful about revealing sources & methods
    • Content, timing, headers, steganography, other covert channels
      – Careful about sending Trojans / viruses / etc. (content)

• Equities: Is protecting sources and methods more important than providing the results to administrators?
A different deception:

- Time picture of the deception - single thread case:
  - Threat is over-the-wire only - low to medium resource (e.g., attack sites)
  - We use a set of collection points to collect data from their locations
    - Try to do it covertly at some collection points?
      - If they detect some, they will be suspicious and a very good deception will be needed to counter perception as very hostile
    - Try to be overt from each site but not excessive - and covert over all
      - Act like something they expect to come past (e.g., a web spider or some such thing - different bus consistent at different sites).
      - If they detect it, they may get curious - if they detect many they may get more concerned at its scope
  - They detect the collection
    - They ignore it, we get the data, they don't know why
    - They attack back / They trace it back / They correlate
      - Make it innocuous to a level of credibility that they won't question.
      - Make it as consistent as possible with the previous pulls
      - Make some more easily found to get burned and find thresholds!
Many collection points used
Each with different thresholds
Each emulating different things
All feeding (eventually) us
As detected/reacted we burn
Redundancy to detect deceptions
More resource intensive
More complex set of emulations

Diagram of a simple decision process we and they might use
Diagram of a different decision process we and they might use.
Deception Processes

Internal Images
Self (system) image - World image - GOAL - they see a web crawler or two
Knowledge, including data, theories, models of people, physics, etc.

Def ne problem / observe:
- Policy
- Situation

Def ne problem solving status: Hypothesize
- Policy met?
- Have tools?

Determine solution options
- investigate?
- attack back?
- refuse service?
- ignore?

Initiate actions / response
- task resources

Associate:
- multiple intelligence pulls?
- intelligence pull and other things
- web crawler?
- other traffic?

Perceive form:
- looks like lots of these things
- looks like known pattern
- looks like lots of different things
- some detected

Perceive Feature:
- lots of requests
- unusual requests
- unusual correlation across requests

Sense:
- many diverse requests arrive

Stimulus (presentation)
Detect
React

Drive:
- send responses
- implement moves
Response (action at a distance)
operator actions

Implement form:
- detail moves OR
- serve / log / warn
- ignore some of them?
- react noticeably to some of them

Implement feature:
- sequence moves

Conditioned behavior
Intuition

Direct:
- order actions

Reflexes
Response (action at a distance)
operator actions

Stimulus (presentation)
Detect
React

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
**Internal Images**

Self (system) image - World image - **DECEPTION GOAL**: they see a diverse set of things  
They believe some are innocuous, others hostile - we calibrate their decision process  
We keep them entertained to the level of resources applies - and continue to get stuff  
Knowledge, including data, theories, models of people, physics, etc.

When they see a pull they notice, they react to it and we either learn about their defenses or continue to pull  
COTS detection and response technologies likely - others possible.

Our goal is to calibrate their responses and cut off return paths before they get to our real combined cause.  
They see lots of stories, each consistent, deep, and credible as can be relative to their resource levels.

---

**Def ne problem / observe:**
- **Policy**: their policy will set their response rules. Generally this will typically be limited to:
  1. shut down service if harm being done or likely soon  
  2. limit service to any given site with throttles  
  3. search out information on you by what you send  
  4. investigate indirectly (over the Internet, etc.)  
  5. attack you to break in and see what is there  
  6. selective deceptive responses
- **Situation**
  
  You look like different things - some they like

---

**Initiate actions / response**
- **task resources**
  They have limited resources.  
  We want to condition their behavior to let us get our data  
  We don't want to appear to be intentionally low consequence  
  We could consume their resources in feints from elsewhere.
  We find their thresholds and stay below them on some pulls  
  We force their responses toward select assets.

---

**Def ne problem solving status**: Hypothesize
- **Policy met? If so, they are done, otherwise, they must act**
- **Have tools? They will use the tools they have (short run)**
- They will hypothesize:
  1. just a crawler - ignore unless loud or painful  
  2. somewhat suspicious - investigate  
  3. highly suspicious - attack back / deny service / deception  

**Determine solution options**
- **ignore? - we win**
- **investigate? - catch some of us? We win**
- **attack back? - get cut off? We win**
- **refuse service? - we detect threshold(good intel)**
  
  diversify IPs for ongoing pulls, lower rates  
- **deceive? - we detect? We win! - no detect - we lose**

We want to seem diverse enough so they don't catch us all  
We want them to choose ignore, investigate, or attack back.  
We find their thresholds and entertain them.  
We do **NOT** want them to correlate all of us or trace back
Perceive form:
- looks high volume
- looks like a web crawler
- automated responses may be used
- most web sites want to be crawled
- some web sites view as an intrusion

Perceive Feature:
- normal logging of web server use
- automated correlation or presentation
- feature extraction may be done:
  - large volume request sites
  - URLs intended to catch crawlers
  - cookies and URL-based tracking
  - referer fields

- Other features? What are they?

Sense:
- requests arrive along with many other requests of similar sorts
- web server (as an example) allows normal web requests at a finite rate
- reflexes send back results unless rate exceeds, or otherwise directed

Stimulus (presentation)

Conditioned behavior
Can we condition a desired set of responses?
- create expectations of benefits
- create hassles for hostility
- create environmental expectation
- stimulate to find normal response
- detect response thresholds and
- Can we stay below normal correlation thresholds?
- Can we detect thresholds?
- Can we figure out what is an identifiable feature?
- stealth (low speed)
- stealth (diffuse location)
- stealth (search pattern)
- stealth (referer fields)

Detect
Relexes
Reflexes will detect all of these
- Reflexes likely provide what we want unless over-ridden from above. Try to remain in reflexive response mode. Stay low profile and diverse - till detected. Have more easily detected systems so responses can be detected

React

Implement form:
- web crawler response option
- intrusion response option
- detailed move as directed
- warn or correlate

Implement feature:
- sequence moves for response
- prepare warning messages
- prepare correlation message
- alter reflexive responses

Drive:
- fulfill web requests
- implement other responses
- altered reflexive responses

Response (action at a distance)
operator actions
Our (desired) prequil

- Do we know enough? (each step)
  - How do we learn more (generate responses)
- Do they have obvious tracking mechanisms?
  - Cookies / special URLs / referrer requirements?
- Their thresholds of detection (experimental)
  - Try to push them to response levels from some sites and detect responses.
  - Use flexible IPs for cover sites - DHCP via many ISPs?
  - Real cover sites are expensive - but cutouts?
  - How often can we do it? Are they even looking?
- Their decision process (observe/effect)
  - join them undercover / create crawler benefits / do 'media interviews' / elicitation techniques / technical attacks
- Resource limitations (theirs / ours)
• We need to protect the back end from feedback
  – As before...
  – Reflexive controls/Trojans by pushing in deceptions
  – Obvious disassociation (as before)
  – Equities
  – Add:
    • Need to continue flexing identities?
    • For long-term visits, need to vary patterns of usage?
    • Need to keep updating browser deceptions?
    • Level of effort required for the long run?
Computer Deceptions

• Computer 'cognitive' structure
• Deceptions against computer systems
• Computer networks as complex systems
  – With computers
  – With people
  – With infrastructure
Cognitive model of computers

Drivers:
- digital data into internal memory

Hardware:
- signals arrive at interfaces

Stimulus (inputs)

Operating System:
- control resources

Protocol:
- syntax of interaction

Drivers:
- order actions

HW Fault Mechanisms

Memory

Applications:
- specific functional elements for this system's requirements

Embedded languages:
- interpreted languages in applications

Applications:
- prescribe actions

Built-in library responses

Libraries:
- standardized functions

Built-in OS responses

Applications:
- standardized mechanisms

Operating System:
- control resources

Built-in protocol responses

Protocol:
- translate syntax

Drivers:
- actions

Built-in library responses

Operating System:
- control resources

Built-in OS responses

Applications:
- specify actions

Application FSM responses

Applications:
- prescribe actions

Built-in protocol responses

Protocol:
- translate syntax

Drivers:
- digital data into internal memory

Embedded languages:
- request actions of their applications

World - Feedback

Response (outputs)
Cognitive Model of computer deceptions

Drivers:
- concealment?
- simulation

Hardware:
- concealment?
- simulation

Stimulus (inputs)

Embedded languages:
- concealment?
- simulation

Operating System:
- concealment?
- simulation

Applications:
- concealment?
- simulation

Libraries:
- concealment?
- simulation

Protocol:
- concealment?
- simulation

SW Fault Mechanisms

Drivers:
- order actions

Operating System:
- control resources

World - Feedback

Response (outputs)

SW Fault Mechanisms

Operational languages:
- request actions of their applications

Applications:
- Prescribe actions

Libraries:
- Standardized mechanisms

Embedded languages:
- Built-in library responses
- Induced library state/response

Operating System:
- Built-in protocol responses
- Induced protocol state/response

Computers:
- translate syntax

Applications:
- order actions

Hardware:
- actions

Suppressed response

Induced protocol state/response

Built-in library state/response

Induced FSM state/response

Induced OS state/response

Induced language state/response

Embedded language responses

Applications:
- Standardized mechanisms

Operating System:
- control resources

Protocol:
- translate syntax

Applications:
- order actions

Hardware:
- actions

Response (outputs)
Exploiting Computer Cognition

- For all cognitive levels
  - Find techniques to induce or suppress signals
- Itemization by level to follow
  - Based on the all.net security database
<table>
<thead>
<tr>
<th>Hardware</th>
<th>Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoofing</td>
<td>Illegal value insertion</td>
</tr>
<tr>
<td>Cable Cuts</td>
<td>Emergency procedure</td>
</tr>
<tr>
<td>Interference</td>
<td>Error inducement</td>
</tr>
<tr>
<td>Fire</td>
<td>Time-based attacks</td>
</tr>
<tr>
<td>Insertion</td>
<td>Audit supression</td>
</tr>
<tr>
<td>Flood</td>
<td>Multiple error inducement</td>
</tr>
<tr>
<td>Modification</td>
<td>Induced stress failure</td>
</tr>
<tr>
<td>Earth Movement</td>
<td>Electronic interference</td>
</tr>
<tr>
<td>Sympathetic vibration</td>
<td>False updates</td>
</tr>
<tr>
<td>Environmental control</td>
<td>Repair-replace-remove</td>
</tr>
<tr>
<td>Cascade failures</td>
<td>Man-in-the-middle</td>
</tr>
<tr>
<td>System Maintenance</td>
<td>Wire closet attacks</td>
</tr>
<tr>
<td>Environment corrupt</td>
<td>Dependency analysis</td>
</tr>
<tr>
<td>Trojan Horse</td>
<td>Process bypassing</td>
</tr>
<tr>
<td>Device access</td>
<td>Below threshold</td>
</tr>
<tr>
<td>Fictitious people</td>
<td>Hangup hooking</td>
</tr>
<tr>
<td>Simultaneous access</td>
<td>Piggybacking</td>
</tr>
<tr>
<td>Resource denial</td>
<td>Call forwarding fakery</td>
</tr>
<tr>
<td>Implied trust</td>
<td>Collaborative misuse</td>
</tr>
<tr>
<td>Race conditions</td>
<td>Input Overflow</td>
</tr>
<tr>
<td>Driver</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td><strong>Driver</strong></td>
<td><strong>Device access exploitation</strong></td>
</tr>
<tr>
<td><strong>Collaborative Misuse</strong></td>
<td><strong>Input overflow</strong></td>
</tr>
<tr>
<td><strong>Race conditions</strong></td>
<td><strong>Simultaneous access</strong></td>
</tr>
<tr>
<td><strong>System maintenance</strong></td>
<td><strong>Implied trust</strong></td>
</tr>
<tr>
<td><strong>Trojan Horses</strong></td>
<td><strong>Illegal value insertion</strong></td>
</tr>
<tr>
<td><strong>Fictitious People</strong></td>
<td><strong>Privileged program abuse</strong></td>
</tr>
<tr>
<td><strong>Spoofing &amp; Masquerading</strong></td>
<td><strong>Error-induced misoperation</strong></td>
</tr>
<tr>
<td><strong>Insertion in transit</strong></td>
<td><strong>Audit suppression</strong></td>
</tr>
<tr>
<td><strong>Modification in transit</strong></td>
<td><strong>Induced stress failure</strong></td>
</tr>
<tr>
<td><strong>Sympathetic vibration</strong></td>
<td><strong>False updates</strong></td>
</tr>
<tr>
<td><strong>Cascade failures</strong></td>
<td><strong>Network service attacks</strong></td>
</tr>
<tr>
<td><strong>Undocumented functions</strong></td>
<td><strong>DCAs</strong></td>
</tr>
<tr>
<td><strong>Environment corruption</strong></td>
<td><strong>Reflexive control</strong></td>
</tr>
<tr>
<td><strong>Dependency analysis</strong></td>
<td><strong>Piggybacking</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Below-threshold attacks</strong></td>
</tr>
</tbody>
</table>
- System maintenance
- Implied trust
- Trojan Horses
- Emergency procedure
- Fictitious people
- Desynchronization
- Spoofing/masquerading
- Multiple errors
- Insertion
- Viruses
- Modification
- Repairs
- Sympathetic Vibration
- Process bypass
- Cascade failures
- Input overflow
- Unknown function
- Illegal values
- Environment corruption
- Error induction
- Simultaneous access

- Below threshold
- Audit suppression
- Peer relationship
- Induced Stress
- Piggybacking
- False updates
- Collaboration
- DCAs
- Race conditions
- Man-in-the-middle
- Reflexive control
- Replay
- Dependency analysis
- Error analysis
- Interprocess comms
<table>
<thead>
<tr>
<th>Operating System</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Invalid values on calls</td>
</tr>
<tr>
<td>- System maintenance</td>
</tr>
<tr>
<td>- Undocumentedfuncs</td>
</tr>
<tr>
<td>- Trojan horses</td>
</tr>
<tr>
<td>- Environment corrupt</td>
</tr>
<tr>
<td>- Fictitious people</td>
</tr>
<tr>
<td>- Simultaneous access</td>
</tr>
<tr>
<td>- Resource manipulate</td>
</tr>
<tr>
<td>- Implied trust</td>
</tr>
<tr>
<td>- Spoof/masquerade</td>
</tr>
<tr>
<td>- Interrupt sequence</td>
</tr>
<tr>
<td>- Insertion</td>
</tr>
<tr>
<td>- Emergency procedure</td>
</tr>
<tr>
<td>- Modification</td>
</tr>
<tr>
<td>- Desync/time based</td>
</tr>
<tr>
<td>- Sympathetic vibration</td>
</tr>
<tr>
<td>- Multiple errors</td>
</tr>
<tr>
<td>- Cascade failures</td>
</tr>
<tr>
<td>- Viruses</td>
</tr>
<tr>
<td>- Repair replace remove</td>
</tr>
<tr>
<td>- Data diddling</td>
</tr>
<tr>
<td>- Process bypassing</td>
</tr>
<tr>
<td>- Reflexive control</td>
</tr>
<tr>
<td>- Hangup hooking</td>
</tr>
<tr>
<td>- Dependency exploitation</td>
</tr>
<tr>
<td>- Input overflow</td>
</tr>
<tr>
<td>- Interprocess comms</td>
</tr>
<tr>
<td>- Illegal value insertion</td>
</tr>
<tr>
<td>- Below threshold</td>
</tr>
<tr>
<td>- Privileged program abuse</td>
</tr>
<tr>
<td>- Piggybacking</td>
</tr>
<tr>
<td>- Error inducement</td>
</tr>
<tr>
<td>- Collaborative abuse</td>
</tr>
<tr>
<td>- Audit suppression</td>
</tr>
<tr>
<td>- Race conditions</td>
</tr>
<tr>
<td>- Induced stress failures</td>
</tr>
<tr>
<td>- DCAs</td>
</tr>
<tr>
<td>- False updates</td>
</tr>
<tr>
<td>- Error analysis</td>
</tr>
<tr>
<td>Library</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>- Simultaneous access</td>
</tr>
<tr>
<td>- System maintenance</td>
</tr>
<tr>
<td>- Implied trust exploit</td>
</tr>
<tr>
<td>- Trojan horses</td>
</tr>
<tr>
<td>- Emergency procedure</td>
</tr>
<tr>
<td>- Fictitious people</td>
</tr>
<tr>
<td>- Desynchronize / time</td>
</tr>
<tr>
<td>- Spoofing/masquerading</td>
</tr>
<tr>
<td>- Imperfect daemons</td>
</tr>
<tr>
<td>- Insertion</td>
</tr>
<tr>
<td>- Multiple errors</td>
</tr>
<tr>
<td>- Modification</td>
</tr>
<tr>
<td>- Viruses</td>
</tr>
<tr>
<td>- Sympathetic vibration</td>
</tr>
<tr>
<td>- Repair and replace</td>
</tr>
<tr>
<td>- Cascade failures</td>
</tr>
<tr>
<td>- Process bypassing</td>
</tr>
<tr>
<td>- Unknown functions</td>
</tr>
<tr>
<td>- Content-based attacks</td>
</tr>
<tr>
<td>- Environment corrupt</td>
</tr>
<tr>
<td>- Restoration process</td>
</tr>
<tr>
<td>- Dependency analysis</td>
</tr>
<tr>
<td>- Hangup hooking</td>
</tr>
<tr>
<td>- Interprocess comms</td>
</tr>
<tr>
<td>- Input overflow</td>
</tr>
<tr>
<td>- Below-threshold attacks</td>
</tr>
<tr>
<td>- Illegal value insertion</td>
</tr>
<tr>
<td>- Piggybacking</td>
</tr>
<tr>
<td>- Error inducement</td>
</tr>
<tr>
<td>- Collaborative abuse</td>
</tr>
<tr>
<td>- Audit suppression</td>
</tr>
<tr>
<td>- Race conditions</td>
</tr>
<tr>
<td>- Induced stress failures</td>
</tr>
<tr>
<td>- Error analysis</td>
</tr>
<tr>
<td>- Fales updates</td>
</tr>
<tr>
<td>- Reflexive control</td>
</tr>
<tr>
<td>- DCAs</td>
</tr>
</tbody>
</table>
Application

- Salami attacks
- False updates
- Repair replace remove
- DCAs
- Process bypassing
- Replay attacks
- Content-based attack
- Error analysis
- Restoration process
- Reflexive control
- Input overflow
- Dependency analysis
- Illegal value insertion
- Interprocess comms
- Privileges program abuse
- Below threshold attacks
- Error-induced misoperation
- Peer-relationships
- Audit suppression
- Piggybacking
- Induced stress failure
- Collaborative abuse
- Kiting

- Repudiation
- Maintenance
- Environment corruption
- Trojan horses
- Modeling mismatch
- Fictitious people
- Simultaneous access
- Spoof/masquerade
- Implied trust exploits
- Insertion
- Emergency procedures
- Modification
- Desychronize and time
- Sympathetic vibration
- Imperfect daemon exploit
- Cascade failures
- Multiple error inducement
- Invalid values on calls
- Viruses
- Unknown functions
- Data diddling
- Excess privilege exploit
- Race conditions

- False updates
- Repair replace remove
- DCAs
- Process bypassing
- Replay attacks
- Content-based attack
- Error analysis
- Restoration process
- Reflexive control
- Input overflow
- Dependency analysis
- Illegal value insertion
- Interprocess comms
- Privileges program abuse
- Below threshold attacks
- Error-induced misoperation
- Peer-relationships
- Audit suppression
- Piggybacking
- Induced stress failure
- Collaborative abuse
- Kiting
- System maintenance
- Modeling mismatches
- Trojan horses
- Simultaneous access
- Fictitious people
- Implied trust
- Spoofing/masquerade
- Emergency procedure
- Insertion
- Desynchronize / time
- Modification
- Multiple errors
- Sympathetic vibration
- Viruses
- Cascade failure
- Data diddling
- Invalid values on calls
- Repair replace remove
- Unknown functions
- Process bypass
- Environment corruption
- Error analysis
- Content-based attacks
- Reflexive control
- Restoration process
- Dependency exploitation
- Input overflow
- Below threshold
- Illegal value insertion
- Peer relationship
- Error inducement
- Piggybacking
- Audit suppression
- Collaborative abuse
- Induced stress failure
- Race conditions
- False updates
- Kiting
- DCAs
- Salami attacks
- Replay attacks
- Repudiation
A simple computer deception

- UDP echo virus
  - No deception below protocol level
  - Protocol level simulate normal echo function
  - Concealed below thresholds of higher levels

- Simulation produces built-in protocol response of another echo packet
  - Response is itself a UDP echo virus
Deception in Systems of Systems

• Systems interact at their interfaces to the world and nowhere else
  – Physical as well as informational interfaces
  – People can have operations, computers too

• Systems of systems are treated as a set of interrelated network nodes
  – Multiple interfaces linking into internals
A user and a computer
The cell phone listening deception

- **Human**
  - Sense: concealment
    - Normal telephone on indicators not present
  - Form: simulation
    - Battery gets used up, laid off on bad battery
  - Assess: concealment
    - Lack of knowledge leads to dismissal of notion
  - Think: concealment
    - Phone does not help to investigate - trace hard

- **Phone**
  - HW: concealment
    - No UI indication of phone on is given
    - Phone broadcasts
  - App: concealment
    - No audit information
    - Hard to trace actions
The AntiVirus Covert Channel

- **User system:**
  - Library: simulate web traffic
  - Application: simulate error to change firewall
- **User:**
  - Sense: simulate update fail before possibilities tried
  - Form: simulate error generate trained complain
  - Associate: simulate disassociate subsequent traffic from AV
  - Assessment: conceal facts about attempts to bypass security
  - Thought: conceal the true nature of the product's covert channels
- **Firewall:**
  - Protocol: simulate web traffic, then ICMP, UDP, DNS, Telnet, etc.
  - Application: simulate web traffic to fool content checking
- **Firewall administrator:**
  - All levels: concealment and simulation to cause them to fail to block it
Models of Deception

• Three level goal-directed model
  – Specific criteria for retaining or changing cognition levels in target
  – Principles on when and why to change levels

• Mathematical deception model
  – Type I, II, III errors and available information
  – Deception as information affecting information

• Heuristic models
  – Various historical rule sets
  – Rules regarding ordering of deceptions based on risk management principles
Three level model

• Criteria for retaining / changing levels
  – Low level is always preferred when possible
  – Change to mid-level by triggering pattern matching cognitive mechanisms
  – Mid level is preferred when trained or pattern matching response is desired
  – Change to high level by causing cognitive dissonance in the target
  – High level deceptions to induce changes of mind or slow performance on simpler tasks
A Human / Human Organization Model of Deception

Intent:
Objectives / Quality Schedule / Budget

Expectations
Fidelity / Biases
Effort level
Consistency w/ observables

Deception:= induce / suppress enemy signals

High Level

Match
Move
Dissonance

Middle Level

Low Level

World
Mathematical deception model

- **Type I, II, III errors / available information**
  - Type I (omission) from information overload
  - Type II (commission) from sensory deprivation
  - Type III (substitution) from ambiguity

- **Deception H(D) affecting H(M)**
  - Produces log log function
  - Leads to graphical areas and region of efficacy
  - Experimentation required to find regions of efficacy for Type I, II, III errors
Heuristic models

- Various historical rule sets
  - DoD rule sets and historical principles
  - Useful to create a new rule set based on our results with specific bases for their use

- Ordering deceptions and risk management
  - Do high risk (of detection) actions sooner
  - Commit highest consequence resources later

- Development of rule sets for analysis
  - A significant effort would help develop solid rule sets for different conditions
Group behavior

- More ideas - but not all examined yet
  - Personalities and seeming know-how induce missed opportunities and cut off better lines
  - Group members support each other when being critical may be more helpful
  - Belief in perceived success is strengthened
  - Belief in perceived failure induces more stress
  - Much of this is based on personalities
    - We may be able to build better attack groups
    - Technology solutions may improve attack groups
    - Group think acts for the deception in all 3 regions
      - Too little, right amount, too much information
Adding metrics

 Metrics exist for select elements:
  – Known human response characteristics in some areas
  – Times associated with select automated system behaviors
  – For reflex actions reliability statistics exist

 Metrics do not exist for most elements
  – LDRD with UC Davis about to start
  – Looking at mathematics and metrics of deception in information systems
Break Time
Experiments

– To gain insight and hard data
  • On the effectiveness of deception
  • On different deception techniques
  • On how to respond to what situations
  • On what we must keep secret
  • On the impact of deception on target workload

– Experimental methods
  • Standardized open-ended red team exercises
    – Rate and level of progress are the metrics
  • Deception turned on or off with control groups
  • Measure effectiveness on target
Experiments to date

- 3 regions of the deception space:
  - Too much information => Type 1 errors
    - People ignore content they should look for
    - Detection thresholds are set to ignore lots of things
    - Information is devalued in favor of statistics
  - Too little information => Type 2 errors
    - People imagine and make up their own content
    - Detection thresholds are triggered on anything
    - Any information is treated with excessive value
  - Right amount of information => Type 3 errors
    - Consistency is valued and minor flaws forgiven
    - Credibility and confidence are considered high
    - Pattern matching tends to be used
Experimental configuration

– 5 Groups of 3-6 people
– Each group has 4 hours one day a week
– Groups are observed and forms filled out to determine if deceptions were detected.

– Attack room
  • Computers on Internet, internal net, victim net
  • No real distractions to speak of

– Defender room
  • Isolated facility for systems under attack

– Other resources - any they can get
What the target knows about

- What do they know?
  - Nothing of the technology
    - It is mis-associated with some other technology
  - Of the technology but not how it is being used
    - Attacker errors are treated as technology defenses
  - How it is used but not the configuration
    - Increased backtracking but not necessarily progress
  - Full details of configuration
    - Experiments not underway yet
**Experiment 1**

- **Type 3 errors**
  - Right amount of information present
  - Deception is not apparent
  - False target is easily attacked

- **Results:**
  - No team detected the deception
  - Progress against perceived problem was unaffected by the presence of the deception
  - Those who thought they succeeded had high confidence that they had won even though they never touched the real target
Experiment 1 Diagram

- CCD Net
  - 10.0.0.221
  - 10.3.0.1
- SDNS/Logs 10.3.0.1
- Attacker 10.1.0.2
- Attacker 10.1.0.3
- IR/Sniffer
- 10.1.0.*

- R1 10.2.0.1
- R2 10.2.0.2
- Web servers 10.2.0.*

Red Team Net - Hop 1
Experiment 1 Real Setup

Red Team Net - Hop 1

- CCD Net
- V-FlexNet

10.0.0.221
10.3.0.1

Attacker 10.1.0.2
Attacker 10.1.0.3
R1 10.1.129.1

SDNS/Logs 10.1.0.1

IR/Sniffer
F1 10.2.0.1
F2 10.2.0.2

Web servers
Group 1

10.1.0.*
10.2.0.*

R1 10.2.0.1

Web servers
Group 2

10.2.0.*

R2 10.2.0.2
Experiment 1 Attack Graph

Start

Hard
(hint provided)

Find R Box

Easy

Find D Box

Login

Find/Analyze Content

Expand Privileges

Leave Reentry

Login

Find/Analyze Content

Expand Privileges

Leave Reentry
Experiment 1 Results

Progress of attacks in the attack graph over time - Week 1

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
Experiment 2

• Type 2 Errors
  – Not enough information available to attackers
  – Deception is not apparent
  – Objective of attack harder to find and attack

• Results
  – After a while people started to make up magical defenses that didn't exist
  – After lots of effort finding false targets, none believed they had missed the real target
  – The expanded search space induced one team to suspect the very deception in use, but group interaction stopped its pursuit
Experiment 2 Attack Graph

Start

Find R Box

Harder (hint provided)

Find D Box

Login

Find/Analyze Content

Expand Privileges

Leave Reentry

Login

Find/Analyze Content

Expand Privileges

Leave Reentry
Experiment 2 Results

Progress of attacks in the attack graph over time - Week 2

Time in hours

Progress in attack graph
Experiment 3

• Type 1 errors
  – Information overload for resource exhaustion
  – Deception very apparent, hard to miss
  – Hard to differentiate from real targets

• Results
  – Problem complex with deception not on
  – Differentiation slows attack with deception on
  – A cat and mouse game likely to result in escalation on all sides
  – Classic example of fooling machines thus forcing people to build better tools
Experiment 3 Results

Progress of attacks in the attack graph over time - Week 3
Experiment 4

- Type 3 errors - Overrun condition
  - Deception nature/methods known to attackers
  - Deception relatively high fidelity
  - Insider access (overrun) provided

- Results
  - Self-deception observed
  - Deception associated with backtracking
    - False backtracking slowed attackers
    - Uncertainty associated with results caused confusion
  - Inability to differentiate deception from reality
    - Resulting in significant slowing of attack progress
Experiment 4 Diagram

CCD Net

10.0.0.221

10.3.0.1

SDNS/Logs 10.3.0.1

Fake Traffic

Attacker 10.1.0.2

Attacker 10.1.0.3

IR/Sniffer Dazzler

10.0.0.83

Server

10.1.0.*

10.2.46.*

Flex IP User(s)

Red Team Net - Hop 4
Experiment 4 Attack Graph

Dazzled -> Analyze Content -> Identify as dazzle

Ssh into 10.0.0.83

Expand Privileges

Search -> Dazzled -> Analyze Content -> Identify as dazzle

Sniff Traffic

Search -> Dazzled -> Analyze Content -> Identify as dazzle

Find server

Observe interaction

See Dazzle

Analyze Content

Identify as dazzle

Forge traffic

Control 10.2.46.*

Stop rotation

Find file

Extract file

Expand privileges

Find content

Decrypt content
Experiment 4 Results

Progress of attacks in the attack graph over time - Week 4
Experiment 4-6 Results

Progress of attacks in hops 4-6 plotted in parallel
Experiment 4-6 Results

Progress of attacks in hops 4-6 as a single time sequence

Copyright © 2001 - Fred Cohen & Associates - ALL RIGHTS RESERVED
Overall Results

• Confounding Factors
  – No statistically large effect in weeks 1-3
    • Taken as a whole
    • Taken week by week

• Attack Graph
  – Good methodology

• Deception Works
  – Deception always did better
    • Increased time for progress in the attack graph
    • Attackers went down the wrong parts of the graph
    • Attackers believed they succeeded when they failed
Other Effects

- Over time (runs 4-6), deception led to:
  - Reduced expectation of and desire for success
  - Increased attacker attrition
  - Reduced trust in leadership and planning
  - Reduced enjoyment of the attack effort
  - Increased difficulty, distraction, and uncertainty for those being deceived
  - Attack team communications breakdowns
  - Increased backtracking in the attack graph
Another Approach

• Look at attackers rather than defenders
  – Identify attack graphs for attackers
  – Find methods to guide them as desired
  – Apply the methods as an explicit defense

• Requires an error model for attackers
  – Identify error modes
  – Experiment to find ways to induce errors
  – Figure out how errors misdirect the attacker
  – Induce errors to direct behaviors
Basic Attack Graphs

- Characterization of the set of processes by which a system may be attacked
- Typically shown as a flow chart or similar depiction.
- The detailed graph contains many branches and different methods.
- For simplicity we will use an abstraction of attack graphs shown at a 'planning' level
- Demonstrated in "Simulating Cyber Attacks, Defenses, and Consequences"

Example: The generic attack plan shown here
- Models cognitive processes of people and machines at a very simplistic level.
- Consists of a directed set of two recursive sequences with backtracking.
Two Types of Attacker Actions

- **Observe**
  - Observe behavior
    - Traffic, Processes, Files, Responses, etc.

- **Cognition**
  - Detect desired
  - Differentiate desired
  - Analyze for content
  - Assess from world view

- **Cognition (cont.)**
  - Use to guide action
  - General plan:
    - Get in closer (escalate)
    - Exploit current location

- **Act**
  - Stimulate control/content
  - Use previous responses
Information Theoretic Approach

Ideal static attacker cases:
Deception: \((\forall d \in D, \forall r \in R, H(d)=H(r)), (|R| \ll |D|)\)
You can't tell \(r\) from \(d\) and far more \(d\)'s than \(r\)'s
\(\Rightarrow P(r)/P(d) = |R|/|D| = \text{small}\)
\(\Rightarrow\) Certainty of differentiating \(r\) from \(d\) is small

Protection: \(H(r)=0\)
You can't get any information from \(r\)
\(\Rightarrow\) No information is emitted from \(r\)
\(\Rightarrow\) Extremely limited applicability
Error Types

Real Situation

- Miss data
- Miss session
- Miss inconsistency

Perceived Situation

- Make data
- Make session
- Make association
- Make inconsistency
- Misunderstand

Miss/Make Communication
Miss/Make State/Change
Miss/Make Topology/Change
Miss/Make Model/Change
Example IIS Virus Attack Graph

-/+ data
-/+ session
-/+ association
-/+ vulnerability
-/+ depth
-/+ consistency
-/+ inconsistency
-/+ understand

Scan Net  

Find Net  

Find TCP/80  

Overrun buffer  

Try root shell  

Exploit shell  

Attack others  

Success  

Root shell fails  

Overrun buffer fails  

TCP/80 not Found  

P not found  

-/+ data
-/+ session
-/+ association
-/+ vulnerability
-/+ depth
-/+ consistency
-/+ inconsistency
-/+ understand
# IIS Virus Deception Analysis

## Deception Story 1
- I have an IIS server that you hit and it is exploitable.

## Deception Story 2
- I don't have an IIS server / you missed it / the attack failed / root shell failed / no content.

## Deception Strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Action 1</th>
<th>Action 2</th>
<th>Action 3</th>
<th>Action 4</th>
<th>Action 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>+data</td>
<td>Fake IP</td>
<td>Overrun</td>
<td>Try Root</td>
<td>Exploit</td>
<td>Attack</td>
</tr>
<tr>
<td>-data</td>
<td>Hide IP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+session</td>
<td>See self</td>
<td>Fake 80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-session</td>
<td>Hide 80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+assoc</td>
<td>Redirect to false system/network</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-assoc</td>
<td>Find self</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+vuln</td>
<td>Fake IP</td>
<td></td>
<td>Redirect to fake victim</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-vuln</td>
<td>Hide IP</td>
<td>Attack self</td>
<td>Conceal real vulnerable systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+depth</td>
<td>Pre-use condition</td>
<td></td>
<td>Reroute to fake system in depth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-depth</td>
<td>Fake IP</td>
<td></td>
<td>Conceal shell with false prompts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+consist</td>
<td>Fake IP</td>
<td></td>
<td>Create consistent content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-consist</td>
<td>Hide IP</td>
<td></td>
<td>Create intentionally confusing content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+under</td>
<td>Hide IP</td>
<td></td>
<td>Create false content that makes sense</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Deception story: I have an IIS server that you hit and it is exploitable.
- Deception story 2: I don't have an IIS server / you missed it / the attack failed / root shell failed / no content.
The Dynamic Case

A set of state machines: \( \{(I,O,S,I \times S \rightarrow S', I \times S \rightarrow O)\}_1, \ldots \) with some known states and transition mechanisms.

Two cases:
- Deception: Induce or suppress inputs
  => cause or inhibit state transitions
  -> Within states - the static case
- Protection: Suppress inputs
  => cause or inhibit state transitions
  -> Within states - the static case

In order to understand this we need models of the state and state transition mechanisms and we will, no doubt, meet issues of undecidability and NP-completeness as the nature of complexity meets the ability to model
Topologies

- 3 deception topologies:
  - Enveloped
  - Proximate
  - Distant

- Cognitive system limits:
  - Expectation
  - Cognition
  - The Method

- Attack Process
  - Progress (time)
  - Direction
  - Fault types
Strategies

- **Attacker**
  - Goal is proximate
    - Hardest to deceive
    - Most real content
    - Method easiest to apply
  - Enveloped is isolated
    - Easiest to deceive
    - Content controlled
    - Method very hard
  - Distant is powerless
    - Moderate content deception
    - Method works well
    - Little content available

- **Defender**
  - Goal is distant
    - Little content available
    - Deception for warning
    - Method observable
  - Enveloped is OK
    - Total content control
    - Deception easily done
    - Method obvious, hard
  - Proximate is worst
    - Real hard to deceive
    - Real content exposed
    - Method hard to avoid
The timeline below describes how an attack process flows. Over time, a sequence of states and transitions are recorded. States are from the attack graph, transitions are from the cognitive process graph, and computer-induced from computer cognition graph.
Attacker seeks target - via scanning IP address space - NO ERRORS
Attacker sees false target
- MAKE TOPOLOGY - via false responses
Attacker tries to figure out what it is using cognitive process
Attacker believes it is a fake - responses meet expectation of fake - NO ERROR
Attacker seeks target - via scanning IP address space - NO ERRORS
Attacker sees false target
- MAKE TOPOLOGY - via false responses

Seek Target → D/K → Think Fake → Find false target → Differentiate → Think Real → Seek Vulnerabilities

Assume/Believe
Failure - DK - Success

Miss/Make Model/Change
Miss/Make Topology/Change
Miss/Make Communication
Miss/Make State/Change
Attacker tries to differentiate - using expectations
- MAKE TOPOLOGY
Attacker believes target is real
- MAKE TOPOLOGY- meets real target expectations

Seek Target -> D/K

Fail to find real target -> Find false target

Differentiate

Think Fake -> Think Real

Seek Vulnerabilities
Assume/Believe
Failure - DK - Success

Start

Miss/Make Model/Change
Miss/Make Topology/Change
Miss/Make Communication
Miss/Make State/Change
Attacker seeks vulnerability - using known attack scanner
- MAKE COMMUNICATION - deceptions emulate vulnerabilities
Attacker tries to exploit known vulnerability - using script
- MAKE STATE - deceptions emulate state changes
Attacker seeks vulnerability - using known attack scanner
- MAKE TOPOLOGY - deception emulates vulnerabilities
Attacker tries to exploit false vulnerability with custom attack
- MAKE STATE - deception emulates state changes
Attacker tries to exploit false access by seeking content
- MAKE STATE CHANGE - deception emulates content
Deception believed detected
Attacker seeks another target - IP address scanner
- NO ERROR
Seek Target

Assume/Believe
Failure - DK - Success
Seek Vulnerability

Assume/Believe
Failure - DK - Success
Try to enter
Assume/Believe
Failure - DK - Success

Seek Vulnerabilities

Miss/Make Model/Change
Miss/Make Topology/Change
Miss/Make Communication
Miss/Make State/Change

Start

Fail to find false target
Find real target
Differentiate
Think Real

D/K
Think Fake

Fail to find real target
Find false target
Differentiate
Think Real

Seek Target

Try to enter
Assume/Believe
Failure - DK - Success

Find false target
Try to enter
Assume/Believe
Failure - DK - Success
Analyze Target Appearance - using generated expectations - NO ERROR
Target believed to be fake - similar to prior fakes
-MISS TOPOLOGY

Seek Target
D/K

Try to enter
Assume/Believe
Failure - DK - Success

Seek Vulnerabilities
Assume/Believe
Failure - DK - Success

Think Fake
Think Real

Miss/Make Topology/Change
Miss/Make Communication
Miss/Make State/Change

Miss/Make Model/Change

Fail to find false target
Find real target
Differentiate
Think Real

Fail to find real target
Find false target
Differentiate
Think Real

Start
Seek Target - IP address scanner - NO ERROR
Believe real target - based on expectations
- NO ERROR
Seek entry point - using known vulnerability
- MAKE STATE - deceptions emulate state changes
Seek Target

Assume/Believe
Failure - DK - Success

Expand Access

Assume/Believe
Failure - DK - Success

Try to enter

Assume/Believe
Failure - DK - Success

Seek Vulnerabilities

Assume/Believe
Failure - DK - Success

Exploit Access

Start

Assume/Believe
Failure - DK - Success

Try Arbitrary Exploit

Real target

Select Arbitrary Target

No target

Try Arbitrary Exploit

False target

Expand Access

Exploit Access

Assume/Believe
Failure - DK - Success

Deception

Success

Try to enter

Assume/Believe
Failure - DK - Success

Think Fake

Think Fake

Think Fake

Think Real

Think Real

Think Real

Differentiate

Differentiate

Differentiate

Differentiate

Differentiate

Find false target

Find real target

Find real target

Find false target

Fail to find false target

Fail to find real target

Fail to find real target

Fail to find false target
Thank You

http://all.net/  fc@all.net