

# Insider Threat Detection

Using Lightweight Media Forensics

## Cyber Security Division 2012 Principal Investigators' Meeting

October 10, 2012

**Nicole L. Beebe, Ph.D.**  
**Assistant Professor**  
**The University of Texas at San Antonio**  
**Nicole.Beebe@utsa.edu**  
**210-458-8040 (210-269-5647 cell)**

# Outline

- Intro
- Technical Approach
- Milestones, Deliverables, Schedule
- Technology Transition Plan
- Quad Chart

- TTA #4 Insider Threat
  - Focus on detection (vs. prevention or deterrence)
  - Detect anomalous accumulation of data
    - Indication and warning of exfiltration threat
    - Alert to criminal and/or exploitable behavior
- Research Team
  - PI: Simson Garfinkel, Ph.D.
    - Naval Postgraduate School
  - Co-PI: Nicole Beebe, Ph.D.
    - The University of Texas at San Antonio
    - UTSA Co-Investigator, Dr. Daijin Ko, Ph.D.

# Technical Approach

## The Problem

- Indication and warning of exfiltration threat when
  - Information collected IAW access permissions
  - Data collected, transmitted, and deleted quickly
  - No “signature” exists for behavior
- Past detection efforts focused on
  - Monitoring access patterns
  - Defining signatures
- Detection solutions must
  - Be scalable to large, diverse, dislocated organizations
  - Be lightweight and not impact computational operations
  - Accommodate fluidity and variety in workforce

# Technical Approach

## The Solution

- Detect threatening insiders by detecting deviations in their individual storage profile on their workstation
  - Deviations in storage profile
    - Data focused (not quantity/quota)
    - Types (file/data type and content based)
    - Deviations measured from variety of baselines
      - Individual, role-based, department, organization
  - Tunable anomaly alert sensitivity
    - To balance false positive – false negative trade-off
    - To increase monitoring on individuals or units

# Technical Approach

## The Solution

- Benefits
  - System agnostic
    - Leverages but does not rely on file system
  - Scalable
    - Uses reliable sampling of stored data
  - Does not rely on policy statement or discovery
  - Does not require Internet connectivity
  - Does not rely on insider threat modeling or “signatures”
  - No intellectual property costs (uses GOTS)

# Technical Approach

## The Solution

- Approach
  - Local, lightweight, covert, secured surveillance agents installed on workstations calculates storage profiles
    - `bulk_extractor`
      - Open-source, demonstrated capability, v1.3 just released
    - Random sampling
      - Of disk data to minimize computational overhead
      - Of interval to minimize anti-forensics efforts
  - Central management console
    - Surveillance agents send line-based profile data for analysis and anomaly detection
    - Statistical cluster analysis detects outliers

# Bulk Extractor Viewer Screen Shot



**Bulk Extractor Viewer**  
File Edit View Help

**Directory demodata**

- ccn.txt
- ccn\_histogram.txt
- domain.txt
- domain\_histogram.txt
- email.txt
- email\_histogram.txt
- orig\_report.txt
- report.txt
- rfc822.txt
- sector\_stats0
- tcp.txt
- tcptcp\_connections.txt
- telephone.txt
- test
- test2
- url.txt**
- url\_histogram.txt
- url\_searches.txt
- url\_services.txt
- wordlist.txt
- \_thread0.stat

**Feature File url.txt**

1058632019	http://ocsp.verisign.com/
1058633093	http://ocsp.verisign.com/
1059471531	http://ocsp.verisign.com/
1073350371	http://ocsp.verisign.com/
1073351445	http://ocsp.verisign.com/
1075512712	http://ocsp.verisign.com/ocsp/status0
1085239609	http://evsecure-ocsp.verisign.com/
1085240121	http://evsecure-ocsp.verisign.com/
1085559131	http://ocsp.verisign.com/
1085560205	http://ocsp.verisign.com/
1086116452	http://ocsp.verisign.com/
1086117526	http://ocsp.verisign.com/
1086119908	http://ocsp.verisign.com/
1101366875	http://ocsp.verisign.com/
1101367949	http://ocsp.verisign.com/
1101370333	http://ocsp.verisign.com/
1101728091	http://ocsp.verisign.com/
1101729165	http://ocsp.verisign.com/
1110415929	http://ocsp.verisign.com/
1111707928	http://ocsp.verisign.com/
1111708843	http://ocsp.verisign.com/
<b>1124239640</b>	http://ocsp.verisign.com/
1124240555	http://ocsp.verisign.com/
1152049252	http://ocsp.verisign.com/
1152050326	http://ocsp.verisign.com/
1152052710	http://ocsp.verisign.com/
1159537764	http://ocsp.verisign.com/
1159538838	http://ocsp.verisign.com/
1160044379	http://ocsp.verisign.com/
1160045453	http://ocsp.verisign.com/
1163804772	http://ocsp.verisign.com/
1163805846	http://ocsp.verisign.com/
1163808230	http://ocsp.verisign.com/
1172457555	http://ocsp.verisign.com/ocsp/status0
1203685092	http://ocsp.verisign.com/
1203686166	http://ocsp.verisign.com/

**Image File 1124239640 (0x43028918) nps-2009-domexusers.raw**

1124239424	/k.....LR.	2f6b1ea4	f7a39aa6	1ac802e1	7f4c52e3
1124239440	..@.....?.....	0e60ec40	1c7eb90d	de3fc7b4	df87bd5f
1124239456	zjl.....G..ls.W	7a6a312e	03998113	a84720ce	31730d57
1124239472	..x43.....h/...	2dcd7834	33951299	12b9de68	2faae6e3
1124239488	...*...!.f...XWou.	c28a8c2a	c38b2187	66bd8358	576f75bf
1124239504	<.4.]...<...T..n	3caa2687	5dca1015	3c9f84ea	54c10a6e
1124239520	...J....." .>'...	c4fec54a	ddb90711	97227cdb	3e27d11e
1124239536	x..l....."....GC...	78ec9f31	c9f1e622	19dbc4b3	47439ala
1124239552	.....^..... . .b...	5fa01e90	e45ef5ee	7cf17dab	62018ff5
1124239568	M... "V.....v....	4d0bded0	2256a895	cdae8876	aeeeba0d
1124239584	...M...h.....?.....	f3e44dd9	a0fb68a0	ael43bb3	87c1bb02
1124239600	.....0...04...+.	03010001	a381db30	81d83034	06082b06
1124239616	.....(0&0\$...+.	01050507	01010428	30263024	06082b06
1124239632	....0...http://o	01050507	30018618	68747470	3a2f2f6f
1124239648	csp.verisign.com	6373702e	76657269	7369676e	2e636f6d
1124239664	0...U.....0...	30120603	551d1301	01ff0408	30060101
1124239680	....0A..U...:080	ff020100	30410603	551d1f04	3a303830
1124239696	6.4.2.0http://cr	36a034a0	32863068	7474703a	2f2f6372
1124239712	l.verisign.com/T	6c2e7665	72697369	676e2e63	6f6d2f54
1124239728	hawteTimestampin	68617774	6554696d	65737461	6d70696e
1124239744	gCA.crl0...U.%...	6743412e	63726c30	13060355	1d25040c
1124239760	0...+.....0...	300a0608	2b060105	05070308	300e0603
1124239776	U.....0\$...	551d0f01	01ff0404	03020106	30240603
1124239792	U.....0...0.1.0...	551d1104	1d301ba4	19301731	15301306
1124239808	..U... TSA2048-1-	03550403	130c5453	41323034	382d312d
1124239824	530...*H.....	3533300d	06092a86	4886f70d	01010505
1124239840	.....Jk...X.D.l.y	00038181	004a6bf9	ea58c244	1c318979
1124239856	..+.....L...Xn.	992b96bf	82ac01d6	1c4ccdb0	8a586edf
1124239872	..)^.....R..G'/	0829a35e	c8ca9313	e704520d	ef47272f
1124239888	..8....N... "b...?!	0038b0e4	c9934e9a	d4226215	f73f3721
1124239904	Op1...8.....U	4f703180	f18b3887	b3e8e897	00fecf55
1124239920	.N\$... "Mz...aA.*...	964e24d2	a9274e7a	aeb76141	f32acee7
1124239936	..^...+...>.....W...	c9d95edd	bb2b853e	b59db5d9	e157ffbe
1124239952	..~.....+...;R...	b4c57ef5	cf0c9ef0	97fe2bd3	3b521blb
1124239968	8'?JO...0.....	3827f73f	4a308203	ff308202	e7a00302
1124239984	.....+.....)....2...	01020210	0de92bf0	d482988	18320509
1124240000	^..v..0...*H.....	5e9a7688	300d0609	2a864886	f70d0101
1124240016	...0S1.0...U....	05050030	53310b30	09060355	04061302
1124240032	U81.0...U...V...	55522117	20150602	55040e12	0e566572

# Technical Approach

## Technical Challenges

- Science: Clustering algorithm selection/development
  - Accuracy and speed trade-off of extant algorithms
  - Develop combinatorial algorithm to improve accuracy
  - Need for automated parameter selection amidst noisy data
  - Feature selection
- Engineering: Visualization component
  - Visual representation of anomalies
  - Facilitate analyst drill-down

# Milestones & Schedule

- Year 1
  - Milestone: Develop and test lightweight media forensics agent for Windows workstations
    - Done: Bulk Extractor v1.3 released
    - To do: Enhancements and sampling engine
  - Milestone: Develop and test optimal clustering algorithm
    - Leverage past research at UTSA
      - Self-organizing neural nets in digital forensics
      - Automated parameter estimation heuristics
      - Precision INdex (PIN) measure for combining classifiers

# Milestones & Schedule

- Year 2
  - Milestone: Develop and test management console tools, data aggregation agent, visualization components on limited scope production network at NPS
    - Currently investigating adaptation of open source network collection and remote management systems
- Year 3
  - Milestone: Final development of outlier detection and visualization technology based on test results and large-scale testing on partner network(s)

# Deliverables & Schedule

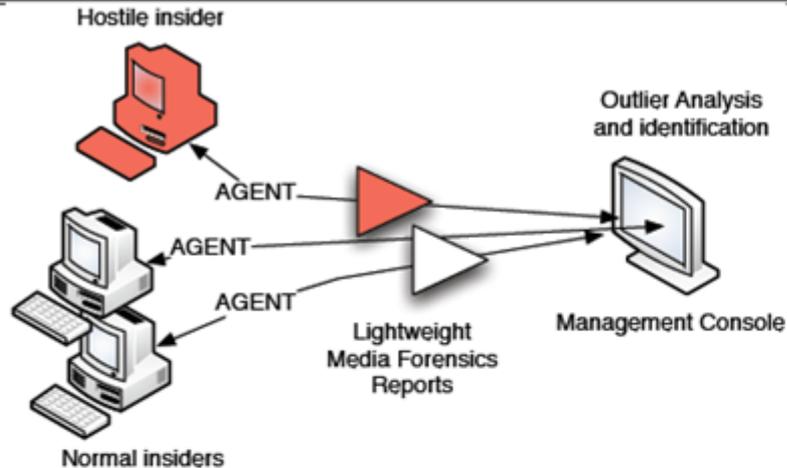
	<b>Knowledge Products (Tech Report / Academic Pub)</b>	<b>Software</b>
Year 1	- Feature Extraction Technique - Anomaly Detection Algorithm	- Workstation Agent (v1)
Year 2	- Visual Component - Small-Scale Network Test	- Workstation Agent (v2) - Server-Side Management Tool (v1)
Year 3	- Large-Scale Network Test	- Workstation Agent (v3 as needed) - Server-Side Management Tool (v2)

- Progress reporting all years
  - Monthly Program Reports
  - Quarterly Status/Financial Reports
  - Quarterly Detailed Technical Reports
  - Annual Program Report
  - Final Technical Report

# Technology Transition Plan

- Open source releases
  - Bulk Extractor (BE)
  - BE sampling engine
  - BE plug-in for clustering engine
  - Clustering engine
  - Visualization front-end for clustering engine
- Create GOTS product that integrates all above, plus
  - Secure transmission of data
  - Data aggregation agent
- To get to pilot: production network tests with injects
  - DETER & PREDICT are not applicable
- PI Garfinkel has extensive experience transitioning technology into open-source, commercial, and GOTS products

## Detecting Threatening Insiders with Lightweight Media Forensics

**Operational Capability:**

- Performance target: Able to scan every workstation in an organization on a regular basis and find systems that have data divergent from historical trends or organizational norms.
- Key Parameters: Types of data tabulated; Kinds of features extracted; Sensitivity of outlier detection.
- Cost of ownership: No IP cost (GOTS-owned technology); deployed via existing management systems; unknown monitoring costs.
- Proposal addresses TTA #4: Insider Threat.

**Proposed Technical Approach**

- Hostile insiders are detected by comparing their storage profile with the storage profile of others in their organization and looking for outliers. This is an advance over the signature-based state-of-the-art and accomplishes the BAA's goal of addressing "scale and diversity of the computing infrastructure," since the approach automatically scales and accommodates different storage profiles at different organizations. It also eliminates the need for policy discovery, since instead of looking for policy violations, it merely looks for people who are storing different kinds of data.
- In the base period a prototype agent, aggregation system, and outlier identification algorithms will be developed.
- We have a command-line forensic tool (*bulk\_extractor*) that extracts all of the features we plan to use for outlier analysis. The tool is currently used by law enforcement.
- Work to date has demonstrated that outlier analysis is a powerful tool for finding hard drives containing large quantities of financially sensitive information.
- Related effort: We continue to develop *bulk\_extractor* for use by law enforcement as a stand-alone forensics tool.

**Schedule, Cost, Deliverables, and Contact Info:**

- **Year 1:** Demonstration of agent and feasibility of outlier analysis.
- **Year 2:** Demonstration of tool on small network with real data and synthetic hostile insiders.
- **Year 3:** Demonstration of tool on one or more enterprise networks.
- **Deliverables:** Software; Research Papers;

Offeror: Simson L. Garfinkel, Naval Postgraduate School  
 Monterey, CA 93950  
 phone: 831.920.8131  
 slgarfin@nps.edu

Administrative POC: Danielle Kuska  
 Monterey, CA 93943-5138  
 831.656.2099 (voice)  
 dkuska@nps.edu