Towards An Adaptable System-based Classification Design for Cyber Identity

PhD Extended Abstract

June 11, 2018

Mary C. (Kay) Michel
School of Computing
Florida Institute of Technology
Melbourne, Florida 32901
michelm1998@my.fit.edu

Michael C. King
School of Computing
Florida Institute of Technology
Melbourne, Florida 32901
michaelking@fit.edu
The Internet...

A dynamic, complex infrastructure that connects us all across the globe.

To understand a cyber identity today, more sophisticated methods are needed.

Identity Analysis Timeline

Cybercrime cases are challenging to solve. Need to organize data in a meaningful way to resolve an identity.
Topic Agenda

• Introduction
• Background
• Initial Design Experiments
• Conclusion and Future Work

My background...
• C.S. PhD Candidate in Computer Science at FIT
• 15yrs. as Chief Software Engineer & R&D P.I.
• 10yrs. researching Cyber, Identity, AR, and VW’s
• Artist and Market Analyst

Dr. Michael King’s background...
• Associate Professor and Research Scientist at FIT
• Director, Applied Research and Innovation
• Program Manager and Research Scientist
Introduction

Novel Classification and Situational Awareness (SA) Approach

• **HYPOTHESIS:** If we take a *holistic approach* of structurally designed common theoretical attributes of information, cybernetic systems, and adaptable DNA, we can repeatedly assess all observed variables numerically within the four dimensions of context, physical, cyber, and human aspects to resolve an identity in feature set template snapshots and eventual time series.

• **QUESTION:** Can a person’s cybernetic DNA fingerprint be structurally designed with biologically inspired robust feature sets of human, physical, and cyber system-based attributes with high entropy and contextually linkable traits?

• **TECHNICAL GOAL:** Holistic adaptive identity feature classification scheme or information architecture
Background of Holistic Identity

• **Superidentity Project**[^2]: our approach is an *adaptive, system-based design* with temporal aspects of binding a person's cyber to physical identity core based on **4 dimensions of context, cyber, physical, and human**

<table>
<thead>
<tr>
<th>Superidentity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyber</td>
</tr>
<tr>
<td>Psychological</td>
</tr>
<tr>
<td>Biographical</td>
</tr>
<tr>
<td>Biological</td>
</tr>
</tbody>
</table>

• **Situation, Barwise and Dempster-Shafer Theory Ontologies**[^3]: our approach *adds cyber aspects*

<table>
<thead>
<tr>
<th>Physical Identity of a Person (Suspect/Criminal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criminal Case</td>
</tr>
<tr>
<td>Situation</td>
</tr>
<tr>
<td>Location</td>
</tr>
<tr>
<td>Biographical</td>
</tr>
<tr>
<td>Behavioral</td>
</tr>
<tr>
<td>Biometric</td>
</tr>
</tbody>
</table>

• Other methods:
  • Multi-Biometric
  • Social Network Connections
  • Psychological, relationships, physiological
Design Aspects

- **Structure** to link attributes across temporal dimensions
- **Common, variant and salient categories** of attributes within dimensions\(^3\)
- **Information and set theory** to mathematically measure feature sets\(^3,6\)
- **Cybernetics theory** to determine a type of control system\(^4\)
- **Bio-inspired adaptive sophisticated design structures** to track mutation/change variables over time\(^5\)
Concept Design

- **Adaptive**, sophisticated linkable structure with types and expression
- **Cybernetic** systems, function, and state
- **4 Dimensions**
- Similarity and variation measures
- Low and high dimensional data
- OSI model paths, links, and identifiers
- Quantitative categorical feature sets
- Consistent models for baselines, cybercriminals, profiles, unique person
- Goal to achieve high match accuracies and rapid identity resolution

Authors: K. Michel, M. King, Florida Institute of Technology

6/11/2018
Initial Design Thought Experiments

Initial experiment results are promising yielding technical contributions below:

- Identity extended cyber-physical feature set classification-based ontology design
- Identity core with temporal states of mapped features, profile, and unique identity evidence
- Contextual cybercrime salient feature sets with aspects of real-world cybercrime cases

Initial experiments were promising in resolving an identity for several types of cybercrime.
Visualization #1: Swatting Identity Theft Sequence Diagram of Evidence[7]

**Context**

1. **Gamer** hires **Tyler Barriss**
2. **Tyler Barriss** in LA starts **VOIP call**
3. **VOIP call app** uses **spoofed # service**
4. **VOIP call threat** is to **Wichita City Hall**
5. **Wichita 911** calls back the phone #
6. **Wichita 911** records **voice conversation**
7. **Caller** says **“he shot his father and has hostages”**
8. **Caller** says his address is **“1033 W. McCormick Street, Wichita, KS”**
9. **Caller** says his **“house is 1-story”**
10. **Wichita 911** dispatches **SWAT team**
11. **Wichita SWAT team** shoots **Andrew Finch**
12. **Andrew Finch** dies of gunshot wounds

**Cyber**

1. **Tyler Barriss** starts **VOIP phone call**
2. **VOIP call app** uses **spoofed # service**
3. **VOIP call is to Gamer Keemstar**
4. **Tyler Barriss** says **“he is SWAuTistic”**
5. **Tyler Barriss** admits **“he Swatted Finch”**
6. Police suspect **Tyler Barriss** is **SWAuTistic**
7. **Wichita police issue a fugitive warrant**

**Physical**

1. **Gamer** hires **Tyler Barriss**
2. **Tyler Barriss** in LA starts **VOIP call**
3. **Wichita 911** dispatches **SWAT team**
4. **Wichita SWAT team** shoots **Andrew Finch**
5. **Wichita 911** calls back the phone #
6. **Wichita 911** records **voice conversation**
7. **Caller** says **“he shot his father and has hostages”**
8. **Caller** says his address is **“1033 W. McCormick Street, Wichita, KS”**
9. **Caller** says his **“house is 1-story”**

**Human**

1. **Tyler Barriss** starts **VOIP phone call**
2. **Tyler Barriss** says **“he is SWAuTistic”**
3. **Tyler Barriss** admits **“he Swatted Finch”**
4. **Wichita police issue a fugitive warrant**

**Swatting**

1. **Gamer** hires **Tyler Barriss**
2. **Tyler Barriss** in LA starts **VOIP call**
3. **VOIP call app** uses **spoofed # service**
4. **VOIP call threat** is to **Wichita City Hall**
5. **Wichita 911** calls back the phone #
6. **Wichita 911** records **voice conversation**
7. **Caller** says **“he shot his father and has hostages”**
8. **Caller** says his address is **“1033 W. McCormick Street, Wichita, KS”**
9. **Caller** says his **“house is 1-story”**

**Alert:** Swatter Impersonation Profile Behavior

**OS / IF’s / versions**

- Application
- Presentation
- Session
- Transport
- NW
- Data Link

**Innocent victim**

**Swatter**

**MATCH ID:**

- **Alias, Face, Voice to a Name**
- **12/29/17**
- **1033 W. McCormick St**
- **Wichita City Hall**
- **1-2 story**
- **L.A.**

**Authors:** K. Michel, M. King, Florida Institute of Technology

6/11/2018
Datasets and Attributes...

**Datasets**

- Current research involves simulation of real-world cybercrime evidence
- Building template models of baseline profiles, cybercriminals, and known persons
- Plan is to use these models with real-world public Social Network (SN) datasets and/or API’s (e.g., Twitter)

**Attributes** that create a unique feature sets

- Salient features in a situation or context
- Categorical (e.g., behavioral)
- Cyber infrastructure, identifiers, aliases
- Physical infrastructure, identifiers, devices
- Person-specific variance
- Linkable features
Conclusion and Future Work

Classification design is promising.
Ready for implementation and effectiveness evaluation:

1. **Finalize cyber identity design** of feature ontology/RDF models, property graphs, inference engines, and the computational framework

2. **Create representative cybercrime use case scenarios** that include identity theft and fraud feature sets

3. **Create experimental framework prototype and Cybernetic DNA fingerprint feature snapshot templates** (e.g., baseline, cybercrimes, profiles, and unique identities)

4. **Perform match and entropy machine learning algorithm method trade** analysis for feature optimization

5. **Conduct use case scenario prototype trial runs** with varying levels of fidelity in simulated data and online data service methods with publicly available datasets and accessible API’s such as Twitter to determine salient identity variables and thresholds for templates → determine signature patterns

6. **Analyze match similarity and variance results** with repeatable experiment trials using feature set evidence

A second set of experiment trials will evaluate the design adaptability and change operators, along with qualitative evaluation of multi-dimensional support for mixed media SA visualizations (e.g., Reality, Augmented Reality, and Virtual Reality) based on the cyber identity classification architecture.
References


Thank you!

Questions?
Back-up Reference Slides
**IoT Botmaster Use Case: How did the police locate the suspect?**

**Context:**
1. **Malware ELF Linux Executable SW Mirai** scans Internet hosts for open port 23 (telnet)
2. **Malware ELF Linux Executable SW Mirai loaders** enters weak password attacks on IoT devices with Busybox OS shell & specific CPU type
3. **IoT device** becomes a Botnet listener connected to C2 Botnet server

**ALERT:**
1. Botmaster Anna-Senpai alias confesses on hacker sites
2. Botmaster Anna-Senpai alias is Paras Jha, Rutgers Student
3. Mirai IoTs linked to Jha, Josiah White, and Dalton Norman
4. Botmaster charged with Computer Fraud

**GOAL:** To weaponize 100+ hardware devices to DDOS target servers

**Q:** How do authorities link IoT Botnet attacks to Botmaster?

Alaska FBI identified infected IoT devices IP address, then subpoenaed GCI state telecom to link a name and physical location. Then, they confirmed these owners did not permit the malware on their devices. They tracked connections back to Mirai C2 server, then court ordered server account owner emails and cell phone # that link to names. Some connections were cloaked in France TOR exit nodes.

**Challenges:**
- Tracking game servers, TOR, attacks, financial records, and alias via packet capture (PCAP) analysis
- **IoT Botmaster Motive:**
  - Initially DDOS to slow Mindcraft competitor servers so players go to theirs for $, then to make $ for fixing the website server attacks

**Results:**
- Attack Signatures based on rate (e.g., Tbps)
  - ~600mbps -1Tbps UDP Flood Port 53 (Krebs)
  - ~1 terabit (OVH)
  - ~1.2 terabit (DYN)
  - 130M SYN per second
  - 450M HTTP per second from 175K IP Addr
  - 4M ACK
  - Target entire IP Range
  - Scanner behavior
  - Malware in flash volatile memory

**ALERT:**
- IoT Botmaster profile behavior - attack Code has Minecraft game server links

**Match ID:**
- Alias, Face, NW IP logs to Email, cell #, Name

Authors: K. Michel, M. King, Florida Institute of Technology