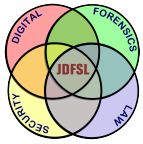

THE JOURNAL OF DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 9 | Number 1

Article 1

2014

Idiographic Digital Profiling: Behavioral Analysis Based On Digital Forensics

Chad M. Steel
George Mason University

IDIOGRAPHIC DIGITAL PROFILING: BEHAVIORAL ANALYSIS BASED ON DIGITAL FORENSICS

Chad M. Steel
George Mason University
P.O. Box 6136
McLean, Virginia 22102
csteel@gmu.edu

ABSTRACT

Idiographic digital profiling (IDP) is the application of behavioral analysis to the field of digital forensics. Previous work in this field takes a nomothetic approach to behavioral analysis by attempting to understand the aggregate behaviors of cybercriminals. This work is the first to take an idiographic approach by examining a particular subject's digital footprints for immediate use in an ongoing investigation. IDP provides a framework for investigators to analyze digital behavioral evidence for the purposes of case planning, subject identification, lead generation, obtaining and executing warrants, and prosecuting offenders.

Keywords: digital profiling, behavioral analysis, forensic psychology

1. INTRODUCTION

Behavioral analysis, once the exclusive domain of the Federal Bureau of Investigation's profilers, had turned into a mainstream area of scientific study. Originally focused on violent offenders, behavioral analysis utilizes concepts like motive, modus operandi, signature behaviors, offender typologies and victim profiles to better investigate criminal activity, understand offender motivations, link criminal acts, and target demographics for prevention efforts.

Digital behavioral analysis is a relatively new field that applies the concepts of traditional behavioral analysis to the digital footprints of criminals. The crimes analyzed can be digital crimes, or those that are digitally facilitated through researching, planning, communicating, documenting, or otherwise enabling criminal activity. Some preliminary work was done in this field by applying a traditional criminological approach to cybercrime. Grabosky proposed a criminological approach to computer crime, providing a categorization of computer-specific offenses (Grabosky, 2000).

The development of typologies and taxonomies of cybercriminals has also been proposed. Krone proposed a typology for a specific type of computer criminal–the child pornographer (Krone, 2004). Nykodym, et al. (2005) proposed a similar typology for insider cybercriminals. Rogers (2010) detailed a taxonomy relevant to hackers that included most traditional cybercrimes including virus writing, hacking, and professional criminals. (Rogers, A two-dimensional circumplex approach to the development of a hacker taxonomy, 2006). Rogers (2010) further applied the concept of social learning theory and moral disengagement toward furthering the understanding of cybercriminal behavior.

Victimology has been studied in several areas of digital crime. Online fraud and how victims are selected was studied as part of a Microsoft study on Nigerian 419 scammers (Herley, 2012). Similarly, Ngo and Paternoster (2011) looked at victim profiles in general across multiple types of cybercrime.

Finally, profiles of user behavior on computers have been researched. In Digital Profiling: A Computer Forensics Approach and Digital Scene of Crime: Technique of Profiling Users, Colombini and Collella (2013) develop a set-theoretic approach to building a usage profile of an individual on a device for the purposes of linking profiles across devices (Colombini, Colella, & Italian Army, 2012).

Most of the prior art takes a nomothetic approach to behavioral analysis by attempting to understand the aggregate behaviors of cybercriminals. This work is the first to take an idiographic approach to digital profiling by examining a particular subject's Internet activities and electronic media for the purposes of

using digital footprints left behind for immediate use in an ongoing investigation.

2. GOALS OF IDIOGRAPHIC DIGITAL PROFILES

Building a profile of a subject in a criminal investigation can be used to provide probable cause for and facilitate the execution of search warrants, assist in subject interviews, link criminal activity, and provide additional case leads. An informative example can be found in the criminal complaint filed against Ross William Ulbricht, aka "The Dread Pirate Roberts", the alleged mastermind behind Silk Road, the Darknet service that facilitated the sale of illegal drugs and banned items over TOR. Silk Road was estimated to have over one billion dollars (US) in annual revenue. Some of the key profile findings that assisted in tracking Ulbricht and obtaining a warrant for his arrest include the following:

The first mention of Silk Road was on www.shroomery.org by the user "altoid", appearing to be a veiled advertisement for the service and provide pointers on how to find it. "Altoid" only posted one message to the site, and directed users to the blog silkroad420.wordpress.com, which was started 4 days earlier by an anonymous TOR user.

Two days later, a user with the name "altoid" posted another advertisement with similar wording for a "heroin store" on bitcointalk.org and pointed users to the same blog.

Eight months later, the user "altoid" posted another message to the bitcointalk.org board looking for an "IT pro". The post requested the user respond to rossulbricht@gmail.com.

The Google account was registered to a Ross Ulbricht. The picture on his Google+ account was the same as a Ross Ulbricht that had registered a LinkedIn account. The LinkedIn account listed Ulbricht as being 29 years old, with a BS in physics from the University of Texas and attendance at a graduate program at the University of Pennsylvania in Materials Science and Engineering. Ulbricht stated in his profile that he was now involved in an "economic simulation" of living in a "world without the systemic use of force" by "institutions and governments.

Ulbricht's Google+ profile contained a link to videos on mises.org. The site had a user profile for Ross Ulbricht with a picture that matched his Google+ and LinkedIn pictures.

The Dread Pirate Roberts contain a link to mises.org in his signature on Silk Road postings. The Dread Pirate Roberts regularly posted using a Pacific Standard Time (PST) time code.

IP address logs showed logins to the Silk Road website from an administrator at an Internet cade near Ulbricht's home in San Francisco.

The logins to Ulbricht's Google account occurred from the home of a friend of Ulbricht's. Ulbricht and his friend posted YouTube videos confirming they lived together.

Ulbricht logged on to the site Stack Overflow with his Google account information and asked "How can I connect a Tor hidden service using curl in php?" One minute after posting, Ulbricht changed his Stack Overflow name from "Ross Ulbricht" to "frosty" and his registered email to "frosty@frosty.com". The SSH key on the Silk Road server was frosty@frosty.com.

The special agents investigating Ulbricht eventually tracked a shipment of fake identity documents that he solicited as the Dread Pirates Roberts to his home, and used the above information to tie him to the illicit Silk Road marketplace. The FBI seized the Silk Road web servers on 2 October 2013 (United States Government, 2013).

The Silk Road forensics work highlights some of the key elements of creating a digital profile for an originally unknown offender. Their investigators found key identifiers associated with the crime, linked the anonymous identifiers to sites that had real name identities, obtained information on the technical expertise and social interactions of the subject, and used IP address geolocation to tie activity in the virtual world to a physical address. This highlights several of the goals in developing an idiographic digital profile:

- Cross-site Tracking. Tracing an individual's actions across multiple sites through their use of common phrases, signatures, or usernames can open up previously unknown leads. Creating a list of relevant sites can also generate a list of locations where the subject's passwords can be obtained more easily in the event strong encryption is

encountered locally, given likely password reuse (Gaw & Felten, 2006).

- Identifying an Anonymous Subject. Anonymous users are likely to break discipline and inadvertently use real information (or a real location) at some point, creating an avenue for identification. Using the cross-site tracking information and legal processes (e.g., subpoenas), a user's true identity can be uncovered.
- Mapping a Criminal Enterprise. The skills and sophistication of a subject can identify their role in a criminal enterprise, ranging from head boss to technical advisor to hired gun. Skills can be criminally oriented, such as building IEDs or hacking, or legitimate skills that can support criminal activity, such as financial or coding expertise.
- Enumerating Associates. Understanding the social network that a subject engages with is helpful in targeting underlings or peers for initial investigative action. By prosecuting other subjects lower down the food chain, investigators can work upward (or sideways, in the case of peers) to the prime subject. Having an assessment of the sociability of the subject can also assist in decision making regarding the likely efficacy of consensual monitoring or account takeover actions.
- Obtaining and Executing a Warrant. The material gathered during the creation of the profile can help link the subject's activities to assist in obtaining probable cause for a search and/or arrest warrant. Obtaining information on the countermeasures deployed by the subject (in the Silk Road case, deletion of information on a VPN server) can help in planning the execution to avoid unintentional or deliberate data destruction.
- Providing Subject Interview Insights. Understanding the motivation and mindset of a subject can assist investigators in theme development for an interview. Additionally, being able to assess the technical skills of the subject provides a barometer to determine if an individual is being deceptive regarding those skills, and allows investigators to have the requisite skills available to assist the interviewers.

The ultimate goal of the proposed framework is to organize digital intelligence regarding a subject into a timely, actionable profile.

3. DIGITAL PROFILE FRAMEWORK

The proposed digital profile framework is broken up into two sections, digital biographical information and a multi-axis competency/affinity profile. The digital biographical information consists of identifiers, websites, signatures, usernames, passwords, and other information that can provide a pattern of usage for a subject. It can also include real life biographical data if that information is known. The profile axes are both quantitative and qualitative–they evaluate the subject's abilities in four areas: technical ability, countermeasures, sociability, and domain ability. Both sections of the profile should be considered dynamic and should be revised as more information is obtained about the subject.

3.1 Digital Biography

The digital biography serves as a tracking mechanism for all currently known (and suspected) information about the subject's Internet activities. The search for information should be iterative–identifying a unique, new username might trigger a Google search for permutations of that same username. Similarly, the identification of a signature in a web forum posting may trigger a search for that same signature, leading to additional usernames on a different forum. The information included can be considered probabilistic until confirmed through independent corroboration.

Generally, a single email address or message posting is the starting point for gathering information. That identifier is then searched for and the relevant, resulting pages are subpoenaed to obtain subscriber information, with any additional identifiers taken from the returns. That information is then collected, and the process is repeated iteratively until all leads are exhausted (Compton & Hamilton, 2011). Information may be obtained directly via subpoena or through a Mutual Legal Assistance Treaty (MLAT), but not all information is likely to be located with providers that are accessible through these mechanisms and some leads may not be able

to be fully explored. The information that should be included in the biographical section includes the following:

- **Identifiers**. Any usernames/email addresses/handles used on any websites are useful in tracking the activities of a subject. The more obscure the username, the easier it is to search for and individuate. When subpoenaing information from providers, any subscriber information, IP addresses that accessed the site under that subscriber's identity, passwords (if available), and answers to recovery questions should be obtained.
- **Passwords**. Because subjects are likely to reuse passwords, any passwords available from sites that do not store hashes (or store non-salted hashes that can be attacked) should be obtained. Subjects may re-use those same passwords on harder-to-break drive encryption like TrueCrypt or PGP, or may use permutations of a previous password. Personal information, including other passwords, can be used to create a custom attack dictionary for tools like AccessData's DNA or the Passware suite.
- **Sites Visited**. Each of the sites visited by the subject can be cross-searched for all of the other identifiers found and the results can be monitored on a go-forward basis. The types of sites visited may provide insight into interests or hobbies, technical competencies, or social contacts that are helpful in building a profile. The investigator should request the web logs of any accesses from the same IP addresses at identified sites. These may include referrer information that links to other sites used by the subject, or browser string history that will provide details about the subject's web access methods.
- **IP Addresses**. IP addresses used by the subject can be obtained based on the web logs from all of the identified sites as noted above, and through subpoenas to the subject's residential Internet Service Provider. The investigator should also search for all IP addresses in Google (some sites leave web logs or similar tracking mechanisms viewable). Depending on the circumstances, investigators can request a trap/trace on any IP addresses of interest, and may want to consider subpoenas to the major search providers for additional activity from those IP addresses. All IP addresses identified should have the date and time noted for later correlation through device forensics.
- **Locations**. Any physical locations mentioned by the subject or associated with the subject (through IP geolocation, for example) should be collected. Posting times (and time zone information) should be collected as well for future use in tracking the subject's movements and determining the subject's current location. Codepages used and browser languages in request strings, if logged, can assist in country-of-origin checks.
- **Associates**. The identifiers of all of the subject's associates, from contacts on social networking sites to individuals using the same IP addresses, should be collected and retained. The decision on whether or not to build a profile on known associates will be an investigation specific decision based on resource availability.

The biographical information can be correlated with any non-digital information acquired from commercial and governmental sources. In the United States, this includes law enforcement databases like the Federal Bureau of Investigation's National Crime Information Center and commercial aggregators like Choicepoint, TLO, and Lexis-Nexis. The non-digital information can be iteratively combined with the digital information until all reasonable leads have been followed.

3.2 Affinity/Competency Axes

As noted above, psychographic information about an offender obtained through digital forensics is used to create a multi-axis profile. The technical ability axis covers a subject's technical skill, as well as their adoption of new technologies (technophilia). The countermeasures axis looks at the subject's use of protective measures both before and after criminal activity. The sociability axis looks at a subject's social interactions, both online and offline. The domain ability axis evaluates the subject's criminally relevant skillset, generally with the help

of a domain expert. While each axis can be quantified, which may be helpful in multi-offender conspiracies when deciding which subject to target, they are more useful as qualitative measures in investigative planning, developing interview themes, and performing investigative actions.

3.2.1 Technical Ability

Technical ability, for the purposes of profiling, consists of a subject's expertise with digital technologies, as opposed to other technical skills (e.g., engine repair). There are two distinct subareas that are of interest in the investigative profile– general expertise and the adoption of new technologies.

General computer literacy can be difficult to assess, even through direct testing. Self-assessment has been shown to be inaccurate (Merritt, Smith, & Renzo, 2005), and the assessor needs to have an equal or greater level of literacy than the subject to adequately evaluate their skills. As such, it is invaluable to utilize digital forensics specialists in making this assessment. While investigators may encounter subjects who have a deep expertise in a narrow area of computing (e.g., printer repair), this is atypical and can be accounted for by noting the discrepancy in skills as part of the profile. Subjects can be grouped into five categories based on their general computer abilities.

3.2.1.1 Functionally Illiterate. These subjects are not likely to make use of digital technologies. They will have little to no online footprint beyond a single email account, and if they do utilize a computer it is to perform a specific task, such as checking email, that they have learned through rote memorization. If they have a cell phone at all, it is likely to be a feature phone and used solely for voice calls. They are not likely to own or use digital cameras, tablets, or other high tech gear.

The functionally illiterate subject will resist adopting new technologies unless provided a use case that makes it impractical to avoid. There will likely be minimal digital evidence to examine with these subjects, though the use of older technologies may be more common due to their comfort level and memorization-based understanding. In general, individuals that are functionally illiterate don't require a digital profile.

3.2.1.2 Casual User. The casual user is the most common subject encountered. These subjects grew up using digital technology or acquired skills and built proficiency through extensive work or personal use. They will use technologies that they are comfortable with, and will adopt new technologies as they become more commonplace.

The casual user may have gaps in their knowledge, but will know how to conduct Internet searches, install software, send emails and instant messages, and take pictures with their smartphone and send them via MMS. The casual user does not understand nor seeks to understand the science behind most of what they do, does not read technical blogs, and is not interested in technology for technology's sake. The amount of digital material that the casual user possesses is going to be a factor of their discretionary income and their need to keep up with the Joneses. They will regularly upgrade their cell phones every two years, will own a tablet and a laptop, and may have a digital camera lying around. The casual user is not likely to have multiple hard drives or extensive amounts of external storage beyond what they use for backup.

The casual user may have an expansive online footprint. Extensive use of social media and the presence of a small number of email accounts are not uncommon, and are bounded by the sociability axis rather than technological understanding. They are likely to use a single search engine, and may regularly visit web locations based on their non-computing interests. The casual user has no problems ordering goods from Amazon, watching Netflix on their Xbox, or doing banking online.

3.2.1.3 Power User. The power user is differentiated by a love of technology, but does not have a formal background in computer science or computer engineering. The power user is very likely to utilize preventative measures (see countermeasures below) without a deep understanding of how to deploy them. They may utilize software like TOR out of curiosity, and then abandon it shortly thereafter. They will have multiple email accounts and an extensive online presence. The power user is more likely than the casual user to adopt multiple online personas, and may use different personas for different actions.

Power users understand how technology works together, but are missing many of the foundational

concepts of computing. They know what an IP address is, but do not understand how routing works. They will be able to talk about the features of the latest chipset, but would not be able to build a logic gate. The power user is also likely to overestimate their knowledge base in relation to others.

Power users are very interested in new technologies, and will acquire the latest and greatest toys to play with. The power user will install numerous software packages on their systems. Executing a warrant on the residence of a power user is likely to require extensive time, as they will have multiple devices from most current technological categories.

3.2.1.4 IT Professional. Unlike the power user, the IT professional uses technology as a means to an end. They are likely to have a degree or other formal training in information technology, and are likely to hold certificates in networking or system administration. They may have programming skills, and possess an accurate understanding of the terms and concepts related to technology that they use in conversation.

The IT professional, unlike the power user, is more likely to bring home their knowledge and expertise to professionalize their personal technology usage. They are likely to have a backup strategy, to maintain up-to-date antivirus on their systems, and to employ encryption appropriately.

The IT professional may or may not have an extensive online footprint, depending on their sociability. Their usage of technology sites is more likely to be learning and problem-solving oriented, as opposed to gadget-oriented. Some IT professionals may be technophiles–like the power user they might spend discretionary income on tech toys–but they are more likely to understand concepts like upgrade cycles and not necessarily buy or deploy the first version of a new technology.

3.2.1.5 Computer Scientist. The computer scientist has a deep background in computing, with degrees in computer science or computer engineering (although rare, autodidacts at this level do exist). They possess a deep understanding of computer operations, and can develop their own software and hardware if needed.

While the power user employs technology for its own sake, the computer scientist will be more likely to stay with a technology for which they have a deep understanding. They may continue to use a platform for an extended period, staying with Android phones instead of moving to iOS just because a sleek new device is available. Because they have a strong knowledge investment, they may hold on to older systems longer, but once they do switch they quickly attain a mastery level of the new technology. While power users and IT professionals may have programming skills, the computer scientist has software engineering skills. While a programmer can develop new software, a computer scientist develops new algorithms.

Executing a search warrant on the home of a computer scientist should be done with caution. They are not as likely to have made mistakes in setting up their systems, and may have employed less common (or even homebrewed) protections on their systems.

Determining what level of skill a particular subject is at can be challenging, but there are areas that can assist in the determination, including:

- **Education**. Does the individual possess degrees or certifications in digital technology, or have they attended basic or advanced skills training?
- **Terminology**. In the subject's communications, do they discuss technology and do they use technical terms accurately?
- **Sites Visited**. Are the sites they visit oriented toward gadgets, toward implementation guidance, or toward research? Does the subject post on discussion boards related to technology, and are they asking for guidance or providing it?
- **Device Ownership**. How frequently does the subject purchase new cell phones, tablets, or laptops? What does the subject do with their old equipment?
- **Physical Activities**. Does the subject attend conferences related to information technology or subscribe to professional journals?

The subject's technical ability and financial standing both impact their technophilia, or desire to possess and use new equipment. A subject may spend a large amount of their discretionary income on acquiring the latest technology for social standing reasons as well as technical reasons. Because of this, the

possession of the latest device is not necessarily an indicator of technical ability, but it does show a willingness to adopt new technology.

Subjects who have low technical ability but adopt technology extensively are frequently the best individuals to digitally exploit. They are more likely to incorporate technology into all aspects of their life, including the criminal ones, and their low ability may mean they have not taken adequate protective steps (or implemented them properly if they have).

3.2.2 Countermeasures

Related to but separate from the technical axis is the subject's use of countermeasures. While there is some overlap between the subject's technical ability and their use of basic protections, it is not absolute. The computer scientist may not bother to encrypt their hard drive for performance reasons, while the casual user may have a password and encryption employed on their new iPhone because it was recommended by a friend.

Countermeasures can be grouped into two categories–those that are deployed to prevent detection, and those that are deployed to hamper an investigation. Some technologies, such as encryption, can serve both purposes–a child pornographer might encrypt files that they send to other child pornographers to prevent their email provider from detecting the contraband traversing their network. Similarly, they may encrypt their files at rest to prevent them from being used as evidence against them if they are caught.

Although the use of digital countermeasures by criminals has been well documented for decades, (Denning & Baugh Jr., 1999) they have not been evaluated on a continuum to-date. There are multiple levels of digital countermeasure that can be deployed by criminals detailed below, and each represents a higher degree of protection (and possibly paranoia).

3.2.2.1 Passwords. Passwords have become so ubiquitous that their absence is more of an anomaly than their presence. Despite user education, however, most users will choose poor passwords in the absence of complexity controls. Additionally, users will re-use passwords (or variants of passwords) on multiple sites. Choosing stronger passwords and not re-using passwords show disciplined behavior and more complex countermeasures are likely to be encountered. While password reuse is a boon for investigators when strong encryption is encountered, reuse tends to be inversely proportional to the complexity of the password employed (Florencio & Herley, 2007).

3.2.2.2 Device Sharing. Subjects that share physical space with others, including spouses and roommates, may have common devices. These can include anything from wireless access points to laptops, and may have separate user accounts for each individual. Because sharing generally requires setting up an additional account, the act of not sharing is a low-level countermeasure. Subjects that compartmentalize their criminal activities may share some devices but refuse to allow access to others, potentially making the restricted use devices higher value targets when performing a forensic triage.

3.2.2.3 Network Usage. The conditions under which a subject connects to the Internet can show both their technical knowledge and risk aversion. At home, a reasonable countermeasure would be the use of WPA2, which comes pre-configured on most modern routers. Using a wired-only connection may be a countermeasure (or may indicate the subject is a high-end gamer or using older equipment). An aware subject isn't likely to login to their personal email from a hotel kiosk, but they may use open wireless access points in places with few cameras to connect to the Internet semi-anonymously.

3.2.2.4 Basic Software Protections. Most computers come with at least a trial version of anti-virus and anti-malware software pre-installed. Because automatic updates to the operating system are turned on by default in modern operating systems, patch currency is less of an indicator than it used to be. More technical users may custom configure software firewalls, turn off unnecessary services, or run additional anti-malware software. At the extreme, a subject may run a profiling application to identify new applications or services on their system.

3.2.2.5 Encryption. The use of encryption generally requires the subject to take active steps to install and manage additional software. Subjects can use encryption at-rest, and software including PGP and TrueCrypt can provide encrypted files, encrypted containers, or encrypted drives that cannot be unencrypted by brute force if the subject chooses a strong password. Encrypted containers and

encrypted files are of particular interest in that they indicate selective encryption and can provide pointers to areas of interest. At the high end, a subject may employ encryption for network communications as well in the form of a VPN. A subject that is using open wireless access points and a VPN connection to a third party server is utilizing very high levels of countermeasures.

3.2.2.6 Anonymizers. At the easy end of anonymity, a subject may use In Private modes in their web browsing software. While this prevents the recording locally of activity, it does not provide anonymity to the server. For this, subjects need to use web-based anonymizers to hide their browsing. Similar services are available for email via anonymous remailers. Even more sophisticated is the use of onion routing software like TOR to route traffic through multiple hops before reaching its destination. This provides layers of anonymity that are difficult to trace back, but comes at a speed cost. Subjects using TOR have made a conscious decision to trade usability for protection.

3.2.2.7 Steganography. While steganography is much-hyped, in practical terms it has limited uses as a countermeasure. When communicating covertly, steganography can be used to hide content in plain sight, but encryption is a more general purpose tool to transmit secret messages. As such, steganography identified on a subject's machine is indicative of fear of the presence of a message being found out as opposed to the message itself.

3.2.2.8 Counterforensics. At the highest end of the countermeasure spectrum are counterforensics techniques. These include false flag operations (intentionally fabricating forensics information to frame another individual or entity), cleanup routines that alter logfiles to remove traces of a subject's activities, and destructive wiping which makes logical data irrecoverable for later analysis. The use of counterforensics techniques indicates that there is strong technical knowledge present in either the subject or someone advising the subject, and that the subject places a high value on their criminal activities not being uncovered.

Identifying the countermeasures in use can allow investigators to avoid digital tripwires in serving warrants or seizing devices. Additionally, any digital surveillance can be curtailed for subjects who employ more extreme countermeasures as they are more likely to be vigilant about aberrant connections and processes. Finally, the use of extreme countermeasures by individuals with low technical ability may indicate the involvement of outside expertise.

3.2.3 Sociability

Sociability, or the preference for engaging with others instead of being alone, is a more important measurement for profile development than shyness (an emotional tension when interacting with others). The willingness of an individual to engage in social interactions is a more important factor in deciding how to approach an individual than their internal emotional state when the interaction is occurring. Additionally, for online communications, shyness has been found to have an impact on certain technologies but not others. Shyness is negatively correlated with the number of Facebook friends an individual has (Orr, et al., 2009), but not correlated with email or chat usage (Scealy, Phillips, & Stevenson, 2002). While sociability is of primary use, noting factors related to shyness may explain excessive nervousness or anxiety during the baseline questioning in an interview.

The Cheek and Buss five point sociability scale can be used as a baseline for measuring sociability in the profiling process. While their scale includes self-reported answers to questions like "I'd be unhappy if I were prevented from making many social contacts", the same characteristics can be measured indirectly (albeit with less precision) using features extracted during the digital forensics process (Cheeck & Buss, 1981). The following four features should be reviewed to evaluate the sociability of a subject.

3.2.3.1 Sources of Interaction. The different methods that an individual uses to communicate online can be enumerated. Methods may include but are not limited to social media pages, forums, chat rooms, instant messaging clients, and email. The number of different methods and the number of accounts present for each method can be compared to expected numbers based on the person's age, position, financial status, and technical ability. Additionally, the immediacy and directness of interaction should be considered. Posting to a forum does not involve a real-time conversation, and is generally not to a specific person. Skype chats, however, are real time and are closer to in-person

interactions. More direct, extensive sources of interaction would tend to indicate a higher sociability score.

3.2.3.2 Volume of Interaction. While the total number of accounts the subject has is indicative of their signing up for various services, they may do so to test out an application or for a one-time use (e.g., throwaway email accounts used to register for a questionable website). The number of individuals that a subject interacts with and the frequency of interaction with each individual can provide insight into sociability. This can include email contacts, Facebook friends, or chat room partners. In addition to the number of interactions, longer responses to messages and attempts to prolong conversations by asking questions or engaging on other topics can be seen as markers of high sociability.

3.2.3.3 Responsivity. Individuals with a high sociability are more likely to seek out interaction, and a higher rate of conversations that they initiate (as opposed to respond to) is expected. Additionally, developing a forensic timeline of a subject's usage patterns can show how quickly they interact with others once they begin using a device or service.

3.2.3.4 Interaction Duration. Subjects with higher sociability would be expected to have longer conversations, and more verbose and thoughtful qualitative responses to individual messages. For real-time conversations, the exact duration of interaction can be directly measured based on session time. For offline interactions, the time between the first and last posting by the subject can serve as a long-term communications duration.

Individuals who have a large number of meaningful interactions that show positive sociability are more likely to want to engage during an interview. Additionally, they may make better targets for potential consensual monitoring engagements, and are more likely to have spoken with associates about information that may be meaningful to an investigation. For those with extremely high sociability, investigators may only need to make themselves available online in the proper context and the subject may engage them.

3.2.4 Domain Knowledge

The most difficult factor to qualify (or quantify) is the subject matter expertise of an individual in the criminal conduct of interest. For online crimes, the conduct may be hacking ability, virus writing, or the acquisition of child pornography. Offline expertise could include anything from the ability to break into a house to bomb building. Cross-domain criminal skills can include talents that are applicable to multiple criminal endeavors and include areas ranging from observational skills to social engineering.

Ericsson, et al. (1993) identified 10,000 hours of practice as the defining time to becoming an expert in a variety of fields, ranging from chess to music. Similar work has shown that criminals develop expertise in their specific areas based primarily on experience. Wright, et al. (1995) studied residential burglaries and showed that experienced burglars identify more vulnerabilities in homes than lay persons. Additionally, criminal experience has been shown to develop expertise in the perceptions of violent criminals (Topalli, 2004). In the online world, the value placed on criminal technical skills is evidenced by the purchase of these skills by groups ranging from traditional profit-seeking cybercriminals to terrorists (Radianti, Rich, & Gonzalez, 2009; Warren & Streeter, 2006).

Most of the criminals encountered will have sub-expert skill levels in their domain. This provides an opportunity and a challenge. The opportunity is that, if the investigative team has a true expert available, they will likely be able to accurately assess an individual of lesser skills. The challenge is that, in many criminal domains, a subject with sub-expert level skills can still have a large impact, and the difference between a talented amateur and an expert may not be meaningful in developing a profile.

There are several steps involved in building a technical profile based on a subject's criminal domain knowledge. Identifying a relevant domain, assessing the amount of time the subject has spent in that domain, evaluating the subject's use of language and terminology related to the domain, and determining the subject's standing amongst others in that domain are the key factors in evaluating the subject's criminal expertise.

3.2.4.1 Identify Relevant Domain(s). The criminal domains of interest are generally pre-determined from the type of crime being investigated and determined prior to the technical profile development. In a virus writing case, malware development would be the relevant domain. For a

terrorist attack involving a suicide vest, bomb making would be the relevant domain. Less obvious are the secondary criminal domains that may be relevant. The virus writer may have needed expertise on air gaps in place at a location to write an effective virus, requiring research and surveillance skills. Similarly, the terror group may have needed targeting skills to identify a high impact venue and recruiting skills to identify and enlist individuals to deploy their weapons. For criminal enterprises, all subjects should be assessed against the relevant domains for the enterprise as a whole to determine their role(s) in the organization.

3.2.4.2 Assess Experience in a Domain. Because experience is the key factor in expertise, assessing a subject's prior domain experience is valuable. The duration of the experience, coupled with the amount of time the subject focused on that experience, can be partially measured through digital interactions. The first visit to a website or forum related to the domain, or the first email exchange to mention keywords related to the domain, may point to the initiation of interest in that area. This will become increasingly true going forward with increased adoption of services like Google Mail that allow users to retain correspondence indefinitely.

Following the identification of the initial interest, the percentage of online time spent engaged in a domain can likewise be measured. Activity information is generally readily available through proxy logs, Internet history extracted from seized devices, and trap-and-trace order results. While explicit information on interests can be gleamed from correspondence if available, implicit interest can be identified through time spent on particular web pages and the amount of scrolling done (though these are both difficult to measure forensically) (Claypool, Le, Wased, & Brown, 2001).

With an increase in the usage of computer-based training, including online degree programs, formal evidence of related education to a criminal domain may be available as well. An online master's degree in biochemistry may increase the threat potential of a subject browsing information on chemical warfare, whereas completing a certification program as a locksmith would be relevant in a breaking-and-entering case. Similarly, subjects may have related indicators of relevant education, including memberships in professional organizations that have baseline education and experience as criteria to join. This may be apparent through emails from a professional association or online access to restricted journals in a field.

3.2.4.3 Evaluate a Subject's Use of Terminology. There are generally linguistic clues available in a subject's correspondence as to their level of expertise in a domain. A subject's use of uncommon terms particular to a domain, and their proper use of those terms, are related to their level of domain expertise. An individual talking intelligently about the virtues of Classless Inter-Domain Routing is more likely to have an advanced knowledge of networking than a person that refers to opening their web browser as "clicking on the Internet". The terminology can be identified as part of processing correspondence, and looking at term frequency of the subject's correspondence against the baseline term frequency of others in a conspiracy (or against the general public) can quickly tease out differentiators.

Terminology-based assessments can be performed on web searches as well. Jenkins et al showed quantitative differences in how domain experts search as opposed to non-experts. Domain experts were shown to have a more depth-first approach in their search strategies, and this expertise was able to be differentiated from search ability (Jenkins, Corritore, & Weidenbeck, 2003).

3.2.4.4 Professional Standing. While investigators may think of a profession in terms of legal endeavors, criminals have professions as well. They form groups that rely on specialized experience to obtain compensation, and can have hierarchies within these groups that are meaningful. Because there are no board certifications or elections for hackers, their absolute location and status in the knowledge pantheon cannot be definitively identified. Their relative position can be established, however, by an analysis of their interaction with others in their profession.

The primary method for digitally exploiting social networks for expertise is through the subject's online communications. Link analysis of messaging from multiple sources can quickly allow investigators to identify "hubs" – individuals whose expertise is sought by others and have larger numbers of interconnections with other experts. On a micro-scale, individual communications can be examined

to determine the context of the correspondence. If the ratio of queries within a domain that an individual responds to is higher than the ratio of queries they generate, whether in online forums, email communications, or text messages, they are likely to be regarded as having a higher level of expertise. An even more accurate measure, if the correspondence is available, is messages between others previously identified as experts that reference the subject. Sentiment analysis in linked messages that mention the subject can provide an evaluation of their skills that is unbiased, as opposed to being potentially clouded by the deference that may be shown in direct communications due to non-expertise related hierarchical relationships.

Expertise determinations can help link crimes, eliminate subjects, determine how long a subject has been operating, and ensure the investigative team has the necessary skills to pursue the subject. With increasing adoption of digital education and the breadth of digital communications channels available for forensic exploitation, a subject's expertise can be sufficiently approximated before the need for direct interaction.

4. CONCLUSION

The Silk Road case provided an excellent window into how a digital profile can be used in an investigation. The agents involved did an exemplary job and built a digital biography of the Dread Pirate Roberts that allowed them to link seemingly unrelated accounts and activities that ultimately identified the subject. Additionally, they used affinity and competency evaluations as evidence in the complaint process–several of the statements made by Ulbricht relating to coding and server maintenance were presented as evidence of his technical expertise and domain knowledge to establish that he was involved in the development and running of the site. Ulbricht's use of countermeasures became part of his undoing as well–his purchase of fake identity documents and use of encrypted VPN tunnels helped facilitate his identification and arrest. Finally, agents exploited Ulbricht's sociability in communicating with him when Ulbricht attempted to arrange a murder-for-hire hit on FriendlyChemist, a former Silk Road vendor (United States Government, 2013).

While the agents pursuing Silk Road weren't necessarily using a formal digital profiling methodology, codifying their work and the work of investigators who have faced similar challenges allows for the development of a framework for practical use. The guidance presented in this paper is provided to investigators to assist in creating an idiographic digital behavioral profile in active criminal cases. The profile can be developed iteratively and refined during the course of an investigation. When multiple potential users are involved, as may be the case with judicially authorized data intercepts of Internet traffic (e.g., from a wireless access point), profiling can assist in subject disambiguation. Ultimately, a successful profile will provide immediate value to investigators in case planning, subject identification, lead generation, obtaining and executing warrants, and prosecuting offenders.

REFERENCES

1. Cheeck, J. M., & Buss, A. H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology*, *41*(2), 330.
2. Claypool, M., Le, P., Wased, M., & Brown, D. (2001). Implicit interest indicators. Proceedings of the 6th International Conference on Intelligent User Interfaces, *ACM*, 33-40.
3. Colombini, C., & Colella, A. (2013). Digital profiling: A computer forensics approach. *Availability, Reliability and Security for Business, Enterprise and Health Information Systems*, 330-343.
4. Colombini, C., Colella, A., & Italian Army. (2012). Digital scene of crime: Technique of profiling users. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*.
5. Compton, D., & Hamilton, J. (2011). An examination of the techniques and implications of the crowd-sourced collection of forensic data. Third International Conference on Privacy, Security, Risk and Trust (PASSAT), *IEEE*, 892-895.
6. Denning, D. E., & Baugh Jr., W. E. (1999). Hiding crimes in cyberspace. *Information, Communication & Society*, *2*(3), 251-276.
7. Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the

acquisition of expert performance. *Psychological Review*, *100*(3), 363.
8. Florencio, D., & Herley, C. (2007). A large-scale study of web password habits. Proceedings of the 16th International Conference on World Wide Web, *ACM*, 657-666.
9. Gaw, S., & Felten, E. (2006). Password management strategies for online accounts. Proceedings of the Second Symposium on Usable Privacy and Security, *ACM*, 44-45.
10. Grabosky, P. (2000). Computer crime: A criminological overview. Workshop on Crimes Related to the Computer Network, 10th United Nations Congress on the Prevention of Crime and the Treatment of Offenders. Vienna.
11. Herley, C. (2012). Why do Nigerian Scammers say they are from Nigeria? WEIS.
12. Jenkins, C., Corritore, C. L., & Weidenbeck, S. (2003). Patterns of information seeking on the Web: A qualitative study of domain expertise and Web expertise. *IT & Society*, *1*(3), 64-89.
13. Krone, T. (2004). A typology of online child pornography offending. Australian Institute of Criminology.
14. Merritt, K., Smith, D., & Renzo, J. (2005). An investigation of self-reported computer literacy: Is it reliable. *Issues in Information Systems*, *6*(1), 289-295.
15. Ngo, F. T., & Parternoster, R. (2011). Cybercrime victimization: An examination of Individual and Situational level factors. *International Journal of Cyber Criminology*, *5*(1), 773-793.
16. Nykodym, N., Taylor, R., & Vilela, J. (2005). Criminal profiling and insider cyber crime. *Digital Investigation*, *2*(4), 261-267.
17. Orr, E., Sisic, M., Ross, C., Simmering, M. G., Arseneault, J. M., & Orr, R. R. (2009). The influence of shyness on the use of Facebook in an undergraduate sample. *CyberPsychology & Behavior*, *12*(3), 337-340.
18. Radianti, J., Rich, E., & Gonzalez, J. J. (2009). Vulnerability black markets: Empirical evidence and scenario simulation. 42nd Hawaii International Conference on System Sciences, IEEE, 1-10.
19. Rogers, M. K. (2006). A two-dimensional circumplex approach to the development of a hacker taxonomy. *Digital investigation*, *3*(2), 97-102.
20. Rogers, M. K. (2010). The psyche of cybercriminals: A psycho-Social perspective. In *Cybercrimes*: *A Multidisciplinary Analysis*, 217-235. Springer Berlin Heidelberg.
21. Scealy, M., Phillips, J. G., & Stevenson, R. (2002). Shyness and anxiety as predictors of patterns of Internet usage. *CyberPsychology & Behavior*, *5*(6), 507-515.
22. Topalli, V. (2004). Criminal expertise and offender decision-making: An experimental analysis of how offenders and non-offenders differentially perceive social stimuli. *British Journal of Criminology*, *45*(3), 269-295.
23. United States Government. (2013, September 27). Criminal Complaint. Retrieved on October 11, 2013 from http://www.scribd.com/doc/172773407/Ulbricht-Criminal-Complaint-Silk-Road
24. Warren, P., & Streeter, M. (2006). *Cyber Alert: How the World is Under Attack from a New Form of Crime*. Vision Paperbacks.
25. Wright, R., Logie, R. H., & Decker, S. H. (1995). Criminal expertise and offender decision making: An experimental study of the target selection process in residential burglary. *Journal of Research in Crime and Delinquency*, *32*(1), 39-53.