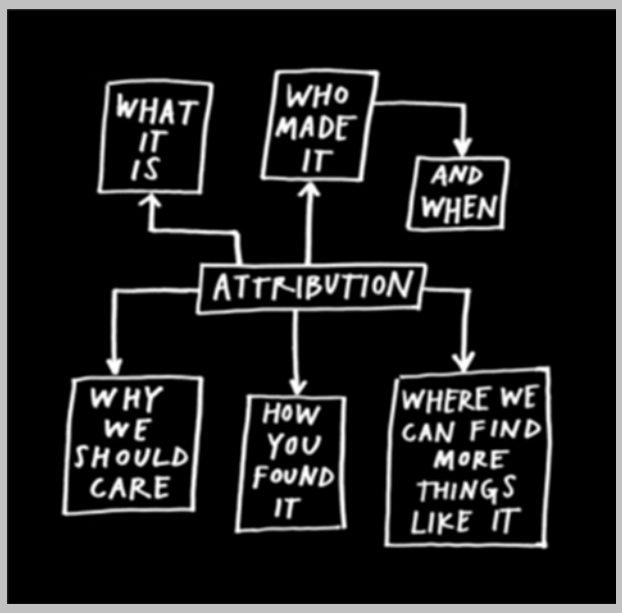
2017 PUBLIC-PRIVATE ANALYTIC EXCHANGE PROGRAM

PHASE II - CYBER ATTRIBUTION USING UNCLASSIFIED DATA



2017 Public-Private Analytic Exchange Program

1 September 2017

PHASE II - CYBER ATTRIBUTION USING UNCLASSIFIED DATA

Abstract

In the shadowy world of cyberespionage, the game of who is to blame can be complicated and fraught with politics, turf battles, national security and geopolitical concerns. Cyber attribution occurs when indicators of compromise (IOCs) and tactics, techniques and procedures (TTPs) from the entire cyber kill chain are associated with an advanced persistent threat or APT group.
Toward the end of the Obama administration, the Department of Homeland Security published a comprehensive list of the tools, techniques and indicators of compromise, called Grizzly Steppe, to out the Russians and their attempts to influence the 2016 presidential election. While the U.S. government has many sources of cyber threat intelligence, deriving from multiple government agencies and private-sector organizations, there is no single approach or framework that extrapolates across domains to derive cyber attribution, definitively and especially as it relates to the unclassified space.

Grizzly Steppe represented the first Joint Analysis Report that publicly identified unclassified technical indicators and attributed them to a nation-state actor. Most often, advanced persistent threat (APT) actor groups are developed independently from multiple sources. Governments, private industry and security companies all use traditional intelligence gathering techniques in their attempts to group malicious cyber activity into functional areas. In other words, if enough cyberattacks use the same indicators of compromise (IOCs) and/or tactics, techniques or procedures (TTPs), the indicators are subsequently assigned or attributed to an associated known actor set. Again, each organization does this differently.

In a 2013 report, Mandiant Corp. publicly exposed APT1 as a Chinese espionage unit, whose activities against U.S. industries were potentially linked to the People's Liberation Army. Subsequently, the FBI released its file on the same Chinese cyber group. This was soon followed up by other federal government agencies acknowledgement that they had been observing this same cyber activity and had developed intelligence on an APT group with the same indicators as APT1 as early as 2002, but had dubbed the group Byzantine Candor and The Comment Group.

Typically, once a pattern of IOCs and TTPs coalesce to the point that they are credited to a unique group, the APT can be attributed. Varied and different approaches are used to develop cyber attribution, most centering upon an exhaustive examination of the IOCs and TTPs to discover a clue that is either aligned with an existing attribution or, in rarer circumstances, evidence of new technical indicators. Since most of the forensic techniques involve similar collateral, security researches will begin their forensic analysis with the malware variants and associated drive-by links of the malicious software, which can reveal deeper or hidden relationships. If a malware sample is already linked to an APT group, then the job is done. New or unseen malware can be attributed with a little extra work. However, it is rarely that simple. Almost all malware "phones"

home to a beacon or command-and-control (C2) server. APT groups will establish a network of these C2 sites within the countries of the target network. They will take several hops to the C2 sites to obscure the ultimate source of their attacks. Security companies and government organizations must therefore engage in "peeling back the layers" to uncover the C2 sites, but without disrupting the site to monitor the traffic to the next hop, and the next, until they can with some level of estimative certainty, focus in on a suspected perpetrator.

Registrant tracking - another technique for attributing involves identifying then monitoring APT actors built infrastructure (say for a spear phishing campaign or malware command/control (C2) hosting) where their registered domains are associated with this hosted infrastructure. An examination of C2 and beacon host names, and email domains, can lead a security researcher to discerning those established patterns. Since much of of this supporting infrastructure is shared across an entity across multiple APT campaigns, there results a confluence of intelligence artifacts associated with the use of those domain registrants: the IP addresses tied to the domain registers and the location and underlying hosting environments used by those actors. Tracking of this registrant information can lead to attribution based on known TTPs of the APT actor. In other instances, companies with mature security programs will employ "threat hunters." These analysts spend their days searching for IOCs and TTPs, proactively, placing "blocks" on internal networks to see if indicators have hit their infrastructure. Hunters often create or participate in groups and communities that share cyber threat intelligence. Threat hunters will pull the same IOCs and TTPs from several sources and validate the data. Yet another technique is that of intelligence sharing. Since many companies lack the depth to independently assign cyber attribution to IOCs and TTPs, other organizations with mature threat intelligence teams, can be engaged to leverage their cyber threat intelligence across larger groups. When malware from spear phishing is discovered, hunters from multiple companies can make the cyber attribution by comparing what they see to other IOC and TTP patterns.

All of that said, the most sophisticated and exhaustive approaches to attribution are often outside the means of most companies, and from the perspective of the government or its intelligence organizations, is usually classified or sensitive. The U.S. government remains compartmentalized in its approach to cybersecurity with no single source of "unassailable truth." This fact, adversely impacts our policy, geopolitical and even military responses. Senior government officials, heads of agencies, corporate executives, investors, and legislators alike share a keen and enduring interest in cyber attribution to support their decision making

The challenge of determining, attributing, deterring, defending against and/or retaliating for such attacks – economically, politically, and/or militarily is driven by this process of “knowing” or “attribution” – which our research shows as much more artifice than science. A variety of theories and approaches abound relative to accurate attribution, characterization and assessment of a cyber

perpetrator or malefactor. De-cloaking the veil of anonymity malicious cyber actors is an extremely difficult task.

Little in the way of formulaic approaches to "attribution" exists, as determined during phase one of our study; even more complicating is the fact that the bulk of the attribution process occurs in the open/unclassified arena, with its vast diversity and dynamic business culture. Prosecutors, law enforcement, legislators, commercial industry, security researchers and the intelligence community all have their specific equities and the effect of this interplay has a notable impact geopolitically and domestically (private/commercial industry impacts). While there exists several models, such as the Diamond Model, that can aid and assist organizations in following a structured approach to navigating the "turbulent" topic of discerning attribution; no clear or broadly applied framework exists.

This multi-disciplinary team was chartered by the Director of National Intelligence (DNI) Public/Private Analyst Engagement Program (AEP) to perform research in this area. The research addressed several key intelligence questions, conducted on-site interviews with subject matter experts, reviewed case studies and hosted panel discussions with cybersecurity experts. Focus areas included the relative importance of attribution to the Public and Private sectors, applicability of certain models (the Diamond Model, ODNIs Framework, etc.) and the state of methodologies/tools appropriate to this endeavor. The research process has concluded with the formulation of a federated framework that will facilitate a more standardized approach to the problem of attribution.

Research Team

A team of experienced cyber personnel was assembled from both government and private industry to address the chosen topic. Each member brought expertise in different areas to support and shape our research.

Name	Organization
Ernest C. (Co-Champion)	Treasury Department
James H.	Aflac
Kyle P.	Ernst and Young (EY)
Christopher P. (Co-Champion)	DHS I&A Cyber Division
Dr. Hector S.	DHS I&A Cyber Division
Steven S.	Lockheed Martin

The team also solicited inputs from various experts in the field (organizations/representatives cited later in the paper) and we would like to acknowledge their assistance.

Methodology

The team began its operations in February 2017, establishing weekly teleconferences to direct its activities and to ensure situational awareness around the overall project plan. The group conducted a research trip in July 2017 to further support of its work. A final product deliverable summarizing those Phase II activities are included in this whitepaper. This final product articulates a "Non-Forensic Attribution Methodology – NFAM" developed to identify perpetrators of unattributed cyber activity against information networks.

Research Trip/Interviews

In support of its research the team attended the BlackHat Cybersecurity Conference and Symposium in July 2017. This site was selected because most of the industry's leaders in cyber-forensics and investigations would be in attendance, thereby facilitating a centralized location to meet with industry experts. To that end, the team met with representatives from:

- FireEye/ISight
- CrowdStrike
- Verizon
- Looking Glass
- ThreatConnect

Research Results

Non-Forensic Attribution Methodology Developed to Identify Perpetrators of Unattributed Cyber Activity against Information Networks

Prepared in collaboration with DHS Office of Intelligence and Analysis (I&A) for the Public-Private Sector – Analyst Engagement Program.

Scope:

This *Reference Aid* provides the findings to an experiment conducted to gauge the validity and reliability of a novel non-forensic attribution methodology (NFAM), meant to complement signature (forensic)-based identification of malicious cyber actors who target information networks. This new attribution methodology represents an alternative use of data sets which can both augment findings from forensic processes and potentially act as an interim substitute if and when assessments made from strict forensic methodologies are unavailable to leadership and network defenders; a situation which has ostensibly appeared to increase over time. Further, this methodology also represents the very first uses of a unique data set within a procedural framework derived from two popular, pre-existing conceptual models namely (1) the Cyber Diamond Model (CDM), and (2) the Confidence Pyramid. The primary audience for this product comprises the security research community.

Key Findings

Upon vetting of the non-forensic attribution methodology (NFAM) in a 4-trial experiment, DHS I&A Cyber Division (referred to hereafter as Cyber Division) found the NFAM could be successfully applied in the identification of cyber actors responsible for unattributed malicious activity conducted against studied information networks. Further, initial findings showed identities of then-masked malicious cyber actors belonging to malicious cyber actors could be accurately assessed at levels of 'likely' and 'very likely' according to the Office of the Director of National Intelligence's (ODNI's) range of estimative likelihoods. Cyber Division also found that using data points, within calculations, representing the level of network compromise achieved by a malicious actor (i.e. Cyber Threat Framework (CTF) stage; see Appendix C) against a victimized network lessened the accuracy of attribution assessments made using this methodology. However, despite this shortcoming, we found that the use of the CTF-derived data points could act as a suitable substitute, in manufacturing a working attribution model for the entities studied, if data on any of three other necessary data categories (i.e. Diamond Model categories of (1) Target; (2) Infrastructure, and (3) Tactics, Techniques and Procedures (TTPs)) were unavailable.

Conduct of Four Trials and Resultant Findings

This experiment was composed of four trials, based on a research methodological guideline used by some graduate-level research programs deemed centers of academic excellence by both the National Security Agency and the Department of Homeland Security. [i, ii] There were four variables which were studied – from which actor profiles were created - to gauge the level of efficacy in the NFAM: (1) the 'Target' of the malicious activity; (2) the tactics, techniques, and procedures or 'TTPs' used by the malicious cyber actor; (3) the 'Infrastructure' leveraged by the malicious cyber actor; and (4) the stage of the ODNI-mandated Cyber Threat Framework (CTF) reached by malicious cyber actors; (*Note: definitions for key terms used in this experiment may be found in Appendix C). Once profiles were created, the data within the nation-state column of entries (i.e. rows of data in the Microsoft Excel spreadsheet recording malicious events) were masked or hidden. Characteristics of the masked actors' profiles were compared to those characteristics of an entry submitted for attribution. The data sets tabulated to form profiles were then compared to a corresponding category in the masked entry to gauge which of the known actors' profiles displayed the highest level of similar activity (either through frequency or impact, depending on the category) in that category. This comparison was done for all categories within an entry. The main outcomes of the four trials were:

- Outcomes for Trial #1, where the data points representing only Target, TTPs, and Infrastructure (and no CTF) were used, demonstrated the highest level (of all trials) of correct attribution assessments against all entities studied; Assessments ranged from 'very likely' to 'almost certainly' on the ODNI's Estimates of Likelihood (EoL) matrix; For more details of the EoL Matrix, see the Instrumentation/Instrument Validity and Reliability section in Appendix A.

- Outcomes for Trial #2, where the CTF data set was added to all data point categories from Trial #1, showed a decrease in attribution accuracy (i.e. efficacy) for most of the major cyber actors in the experiment. One studied entity registered a nominal increase in level of efficacy and remaining 'highly probable' in attribution.

- Outcomes for Trial #3 were not applicable. Infrastructure data points were omitted due to the discovery in prior trials that there were issues regarding question of crossover in IP address usage between entities studied, resulting in these data sets' diminished utility in helping to vet the NFAM optimally. Further, subsequent calculations using data points only from the Target and TTPs categories resulted in too many draws between two or more cyber actors to produce outcomes which could later act as a model for providing attribution (by a sole actor) to unattributed activity. As a result of these preliminary findings, Trial #3 was discontinued.

- Trial #4 was based on the introduction of CTF data as a substitute for Infrastructure data, while once again using Target- and TTPs-related data in calculations. Resultant EoL values for entities a level of 'highly probable' and 'probable' in assessing attribution while one entity's value plummeted to a level showing that an assessment using the NFAM for this data set, would render an assessment which was deemed 'improbable'.

Assumptions

One assumption grounding this study was that data points in the database used were all accurately characterized and inputted, with any amount of inaccuracy being so minimal as not to substantially affect accurate outcomes during calculations. A second associated assumption was that the reporting from which inputs were derived were all accurately captured and reported by those network security professionals from a large variety of enterprises who reported the events. Lastly, there was no literature found showing how much data should be collected (specifically, over what period of time) from which an acceptable attribution model could be derived before efficacy of that model was degraded due to on-going activity by the malicious actor(s). Consequently, the issue of (not) knowing the satisfactory number of inputs for creating a truly representative profile also existed for fixed data sets used in this experiment. Therefore, it was

assumed, for this experiment, that any and all available data in the database leveraged for the time period studied should be used to build the most optimal, representative actor profiles.

Current Challenges to Universal Application of the NFAM

Cyber Division has assumed extrapolation of findings for immediate application to other data sets was limited due mainly to the unique nature of the database used. No other database was known to Cyber Division to exist in the which had accumulated and processed the type and amount of data into a structured and useable format as was housed within the used database. [iii] Consequently, constructing trials representing alternative analyses used to challenge the validity and reliability of the methodology, outside of the controlled environment of the original experiment, was not possible at the time of these trials.

Known, major weaknesses of the NFAM methodology itself included the following: (1) As of this writing, we believed the methodology could only successfully be leveraged to ascribe identity within the same time frame of activity from which profiles were created (i.e. it was not known how well the methodology supported the assumption that 'past is prologue' and, therefore, how well a profile developed from prior activity was applicable to some future time); (2) a lack of automated data processing resulting in semantic (human) understanding of technical terms and the manual input of data points into the database and relevant spreadsheets was likely prone to some unidentifiable level of error, regardless of how inconsequential accumulation of these errors collectively would be to quantitative outcomes; (3) regarding outcomes in the form of numerical values, it was unknown how many inputs at a minimum were required to ensure the efficacy of a value was not skewed by the introduction of a few (errant or anomalous) data points; Lastly, and potentially most destructive to this methodology's efficacy: (4) those database entries for which there was attribution came from entities outside of the database owner meaning unknown fallibility on the part of any of these entities' internal attribution processes may well have been unknowingly magnified in this new methodology.

Outlook

Despite the need for more trials - using more variables - to guide this non-forensic attribution methodology to consistently greater levels of accuracy in outputs, the results of this proof-of-concept experiment have already demonstrated a sufficient level of efficacy as to justify the methodology's immediate use in providing attribution support to network security intelligence customers. Further, the DHS I&A Cyber Division envisions a time (in the near-to-midterm) when the methodology will be so robust that outputs will act as a foundation for indications and

warning capability in support of users with cyber-based equities. Lastly, due to the unique nature of this experiment and the data used, Cyber Division currently lacks the knowledge of the potential external validity of findings or when findings could be incorporated by innumerable network defenders into their respective missions. It is assumed that announcements of continued advancements coupled with computer automation of most, if not all, processes involved will greatly influence both the popularity of and need for the methodology's outputs, particularly in support of cyber-focused policy, acquisition and operational decisions at all affected enterprises.

Appendix A: The Experiment and Trials Conducted

Contents

Background of the Problem

Regarding cybersecurity, attribution of the malicious actors committing cyber-based offenses has become a paramount task for the owners of the victimized networks. The impetus for this prioritization is based on a variety of reasons, but mainly due to hopes of collectively aligning the limited resources (i.e. information on characterization of the actor) of individual enterprises to fight a common adversary who must first be identified. As a subsequent reason prompting the need for attribution, US Government participation is often sought by stakeholders, whether as a leader in a related endeavor or just for material support. Moreover, it is often a characterization of the adversary that must first be made to consider ramifications of actions (both kinetic and non-kinetic) directly supported by the USG, especially when these once-masked actors are later

assessed as associates and even assets of nation-states known to be America's strategic competitors on the global stage.

Despite the differences which exist between the multitude of enterprise networks which have been knowingly victimized, in this work we have assumed the one commonality all victimized network owners share is the majority of activity against their respective networks remains unattributed, a situation which represents a major shortcoming in visibility of threats.

Purpose and Significance of the Study

This experiment served to provide a number of insights. The most important purpose of this experiment was to find to what extent key characteristics of malicious events captured in a database formatted using the aforementioned CDM and CTF could be used for successful attribution of unattributed, observed events. This 'extent' was gauged by the number of positive identifications provided from the total number of entries inputted for evaluation using the model during this experiment. A subsequent purpose of the study was to gauge the influence of using the stage of the CTF (see Appendix C) reached by malicious actors as a variable in calculations.

Due to the unique nature of this experiment, and the lack of knowledge we have on the potential external validity of any findings, one desired outcome was to make findings which could be incorporated immediately by information stakeholders, particularly CISOs and CIOs, depending on the particulars of their respective missions. The significance of such findings would allow disparate entities to use a common lexicon to describe activities and share records of these events which may be affecting other enterprises in a manner which provides participating entities protection of their proprietary information.

Research Question(s)

The following research questions drove the experiment:

1. How effective is a database formatted using a CDM/CTF framework in helping to assign attribution using the three remaining components of the CDM (i.e. Target, Infrastructure, and TTPs) after 'Actor' is omitted?

2. Does the inclusion, within calculations, of the CTF stage reached by a malicious actor increase accuracy of resultant attribution model? If so, to what extent compared to the accuracy without the CTF component?

Theoretical Framework

The theoretical framework undergirding this study was comprised of two concepts. The first concept was the (Cyber) Diamond Model (CDM), used in this experiment to demonstrate how the four key component categories (i.e. (Malicious) Actor, Target, Infrastructure, and Tactics, Techniques and Procedures (TTPs)) could interact in the cyber sphere. The second concept was the Confidence Pyramid which was used in this experiment to show the type (and perceived quality) of evidence within each of the CDM component categories. Based on the Confidence Pyramid, political, technical, and forensic evidence each correspond to confidence levels of low, moderate and high, respectively. (*Note: definitions for key terms used in this experiment may be found in Appendix B of this document, the Cyber Lexicon for the Experiment.) In using the Confidence Pyramid to bolster the effectiveness of the CDM, it was assumed a higher accuracy in analytical assessment would be achieved through knowledge of which pieces of evidence (i.e. data points) potentially represented higher quality inputs for calculation. A graphical representation of the CDM and the Confidence Pyramid, prior to being combined to form a non-forensic attribution conceptual framework, is depicted in Figure 1. The NFAM Conceptual Framework Design (below).

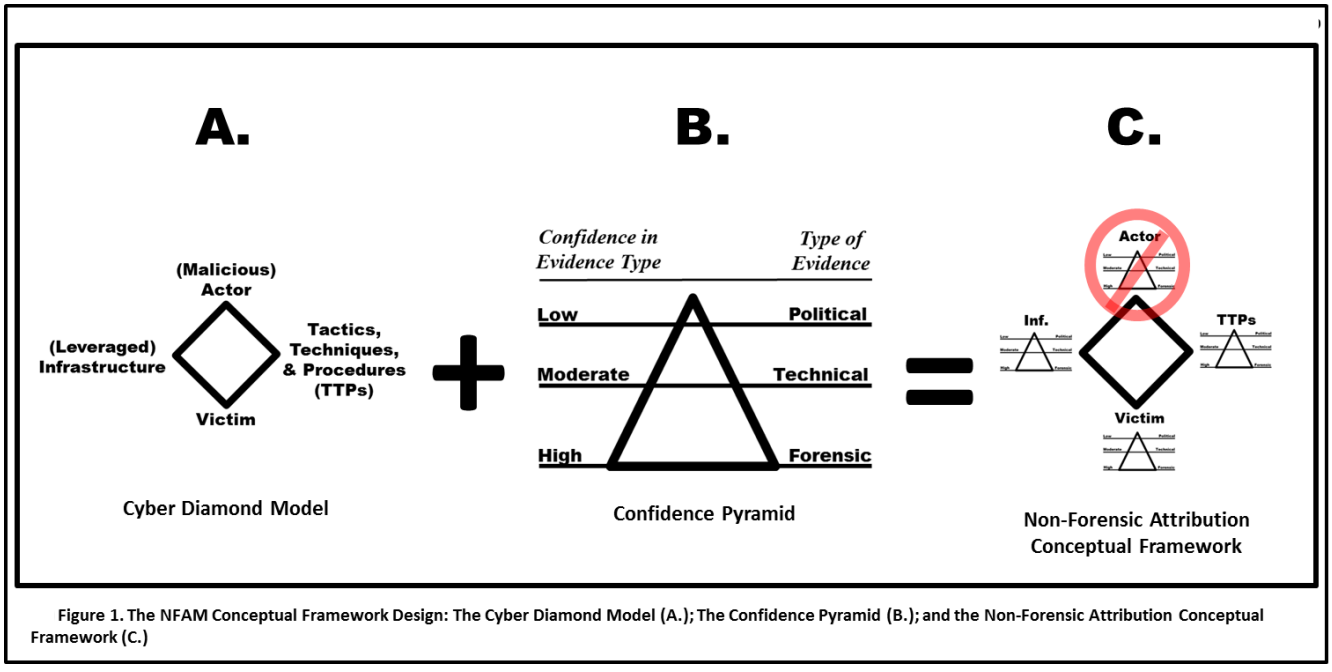


Figure 1. The NFAM Conceptual Framework Design: The Cyber Diamond Model (A.); The Confidence Pyramid (B.); and the Non-Forensic Attribution Conceptual Framework (C.)

Assumptions

One assumption grounding this study was that data points in the used database were all accurately characterized and inputted, with any amount of inaccuracy being so minimal as not to substantially affect outcomes during calculations. A second associated assumption was that the reporting from which inputs into the database were derived were all accurately captured and reported by those network security professionals in the federal government, both in and outside of the IC, who reported the events. Lastly, there was no literature found showing how much data should be collected (specifically, over what period of time) from which an acceptable attribution model could be derived before efficacy of that model was degraded due to on-going activity by the malicious actor(s). Consequently, the issue of (not) knowing the satisfactory number of inputs for creating a truly representative profile also existed for fixed data sets used in this experiment. Therefore, it was assumed, for this experiment, that any and all available data in the database for the time period studied should be used to build the most optimal, representative actor profiles.

Scope, Limitations, and Delimitations

The scope of this experiment was limited to malicious cyber activities observed against entities captured as victims in the used database. However, findings surrounding the efficacy of the methodology used in this study may inform broader audiences and stakeholders in the information network security field, to include military partners and commercial sector partners.

The biggest limitation to (or constraint on) this experiment was the lack of comparable databases available in similar categorization and format to which the data existed. This situation limited the size of the sample frame used in the experiment (which was also the population) in efforts to avoid perceived invalidation of any findings due to control or variable groups sizes being too low (i.e. too many inputs are used to create a profile leaving too few to test against for a believable, statistically-valid outcome). Further compounding this issue was one of the formatting features of the database: the binning of recorded events into the ODNI-mandated CTF is not known to exist in any other database available then to Cyber Division.

Further, the interpretation of verbiage in reporting, deliberation of the stage of the CTF achieved by malicious actors, and the inputting of records into the database is not automated but has all been executed by individuals on a Data Analytics Team. Despite numerous quality control measures in place, the combination of (1) selective reporting (by victims) of events, (2) human interpretation of reports on those events, and (3) the direct (manual) inputting of data

points into the database was believed to collectively introduce at least some (unknown) levels of error which could affect data integrity, regardless of how miniscule.

Delimitations (or self-restraint) were exhibited when data representing only certain entities were chosen for the experiment. In particular, the inclusion of only certain entities was done to preclude the need for cybercriminals to be used within this experiment. If cybercriminals were included, the potential number of criminal entities for whom profiles would have to have been created would have been untenable for the limited resources possessed at the time of this work and under the time constraints for scheduled dissemination of findings.

Research Method and Method Design Appropriateness

The research design of the study was based on guidance provided in the text *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th Edition)*, written by John W. Creswell. This seminal text was deemed appropriate for use in modeling this experiment due to the text's use by some graduate-level research programs managed by US universities which have been deemed centers of academic excellence (CAEs) by both the National Security Agency and the Department of Homeland Security.[iv]

This experiment was composed of four trials that jointly comprised a mixed, predictive/explanatory research methodology.[v] These four trials had to be conducted in sequence as findings in a prior trial acted as a control for the experimentation done in the following trial; it is this workflow which most efficiently provided insights supporting or refuting the two research questions driving this entire experiment. The trials' outcomes were as follows:

- Outcomes for Trial #1, where the data points representing only Target, TTPs, and Infrastructure (and no CTF) were used, demonstrated the highest level (of all trials) of correct attribution assessments against all entities studied; Assessments ranged from 'very likely' to 'almost certainly' on the ODNI's Estimates of Likelihood (EoL) matrix; For more details of the EoL Matrix, see the Instrumentation/Instrument Validity and Reliability section in Appendix A.

- Outcomes for Trial #2, where the CTF data set was added to all data point categories from Trial #1, showed a decrease in attribution accuracy (i.e. efficacy) for most of the major cyber actors in the experiment. One studied entity registered a nominal increase in level of efficacy and remaining 'highly probable' in attribution.

- Outcomes for Trial #3 were not applicable. Infrastructure data points were omitted due to the discovery in prior trials that there were issues regarding question of crossover

in IP address usage between entities studied, resulting in these data sets' diminished utility in helping to vet the NFAM optimally. Further, subsequent calculations using data points only from the Target and TTPs categories resulted in too many draws between two or more cyber actors to produce outcomes which could later act as a model for providing attribution (by a sole actor) to unattributed activity. As a result of these preliminary findings, Trial #3 was discontinued.

- Trial #4 was based on the introduction of CTF data as a substitute for Infrastructure data, while once again using Target- and TTPs-related data in calculations. Resultant EoL values for entities a level of 'highly probable' and 'probable' in assessing attribution while one entity's value plummeted to a level showing that an assessment using the NFAM for this data set, would render an assessment which was deemed 'improbable'.

Infrastructure data points were used only in the first two trials of this experiment to gauge the fullest extent to which the methodology's efficacy could be reached when infrastructure association for a cyber actor was certain. The remaining two trials jointly demonstrate levels of efficacy which can still be reached without leveraging the Infrastructure data, substituting data on the CTF stage reached by a malicious actor in its place.

Although the aforementioned methodology was assumed to be a valid process by which desired insights can be obtained, it was not considered optimal – according to academic rigor – demonstrated by the standards provided in the Creswell text. A more appropriate methodology would have been to create profiles of nation-state actors taken from a first subset of the core data set used, establishing baselines from that first subset, then testing against another second subset or the entirety of the remaining tranche of data. However, two factors precluded this optimal process from being used: (1) a small and/or random set of entries from the same country would not be representative of a nation-state's fullest capabilities and, subsequently, it is assumed such a situation would result in relatively more resultant 'ties' or draws occurring in attribution between nation-state entities; and (2) it was unknown how many inputs were needed to create the most accurate profiles and therefore whether enough entries would have remained (after profiles were established) to perform an experiment against, thus affecting the validity of the findings from such a small sample of the overall data set population (starting at approximately 1400 entries, in total, before the normalizing of data).

The most salient reason for use of the tested methodology was its strong basis in IC tradecraft standard analytical rigor. The construction of the methodology compared any number of competing entities (i.e. hypothesis regarding activity) and processed masked entries (i.e. pieces of evidence) in a manner closer to an analysis of competing hypothesis (ACH) framework

than the optimal academic model of the experiment recommended by Creswell. In using the chosen ACH-based methodology in this experiment, characteristics which were shared by any number of competing malicious actors' profiles were discerned by the corroboration of these common traits to highlight the 'best actor' or potentially by the sheer number of pieces of evidence which tended to support one actor (i.e. hypothesis) above all others considered.

Variables

There were four variables which will be studied in this experiment. They were: (1) the 'Target' of the malicious activity; (2) the tactics, techniques, and procedures or 'TTPs' used by the malicious cyber actor; (3) the 'Infrastructure' leveraged by the malicious cyber actor; and (4) the stage of the ODNI-mandated Cyber Threat Framework (CTF) reached. The Target, as defined for this study victimized USG network, USG data, or USG authorized user. Infrastructure was defined as an IP address associated with and known to be leveraged by the actor. Tactics, techniques and processes (TTPs) were defined as those characterizations of activities reported about victimized networks which have since been processed and binned into CTF 2nd- and 3rd tier categories. The CTF stage reached was defined as the 1st-tier category of interaction with a victimized USG network, USG data, or USG authorized user. More detailed information on CTF tiers and stages may be found in Appendix C (The Office of the Director of National Intelligence's (ODNI) Cyber Threat Framework (CTF)) of this document. All data points characterizing the aforementioned variables must have been captured as part of an entry in the database to be used in this experiment.

Population

The population of the control in this experiment was comprised of all entities, to which attribution of reported cyber-based activity was made, during the years 2015 and 2016, and which were housed within the used database.

Sampling Frame

Due to having no other alternative data sources against which to conduct this experiment, the sampling frame consisted of the very same number of entries forming the population of this experiment (i.e. over 1400 entries).

Regarding homogeneity of the data collected and processed, all of the data points were entered into the database using a unitary format (i.e. one fixed phrase per data field) except under the category of infrastructure. With very few exceptions, values for all entries of the data set used in this experiment within the variables of TTPs, Target, and CTF were populated. However, data points representing Infrastructure were known to be sporadically populated due to the random capture of related information when reports on events were generated by owners of victimized networks. Infrastructure core data points took the form of IP addresses, internet service providers (ISPs), virtual service providers (VSPs), etc. Cyber Division used IP addresses as the sole data type for Infrastructure due to their prevalence compared to other data types mentioned. For the purposes of this experiment, fields not populated with Infrastructure-related data points were left blank. Consequently, the methodology used in this experiment inherently allocated more weight to the other remaining variables, within a submitted entry, when a value for Infrastructure was absent.

Data Collection

The data which was used in this experiment already resided in the leveraged database, an internal database of Cyber Division. The database is comprised of UNCLASSIFIED and sensitive reporting on cyber-based malicious activity. Specifically, the database houses both attributed and unattributed activity reported by victimized enterprises, as well as reports by third party entities about similar targets.

Instrumentation/Instrument Validity and Reliability

According to the Creswell text used to model this experiment, there were three main characterizations for measurement instruments or tools.[vi] They were (1) established instruments used for their historically intended purpose; (2) established instruments (possibly altered and) used for new purposes; and (3) newly developed instruments. This experiment leveraged an established tool and was the impetus for a measuring instrument to be created.

The established tool used for its known functionality was Microsoft EXCEL. EXCEL functions, particularly the spreadsheet, pivot table, and mathematical Macro functions, were used to process and reformat data throughout the four trials of the experiment. This application provided the greatest flexibility in data manipulation as to receive findings as efficiently and quickly as possible. The charting functions of the application were used, after findings were made, to convey points of interest graphically for customers of resultant deliverables.

Despite the abundance of both unclassified and sensitive literature surrounding cyber-based attribution, there was no measurement tool found via literature reviews which appeared to satisfy the requirements for calculating attribution using a data set formatted in the manner used data points were captured. Further, the main requirement of any measurement instrument – used for this particular experiment – had to display both the number of correct assessments regarding attribution made using the methodology as well as the incorrect assessments made. Comparisons between both of these values would allow users of the methodology to gauge the extent inputs and the weights assigned to these inputs could be refined; the desired end state being to achieve the highest number of correct decisions (i.e. 100%) regarding attribution while lowering the number of incorrect decisions. The measurement instrument created for this experiment was called the Hi/Lo Ratio.

A Hi/Lo Ratio (e.g. [X/Y]) represents a 'score' of correct to incorrect assessments, prior to calculation of the more definitive Estimates of Likelihood (EoL) value. The Hi/Lo Ratio is calculated by identifying the number of missed or incorrect assessments made in a data set (i.e. 'Y' or the right side of the ratio) for an actor when the methodology had successfully attributed 100% of what was known to be activity by the actor (i.e. 'X' or the left side of the ratio). The EoL value is calculated simply by dividing the X value by the sum of the X and Y values. This resultant percentage of correct assessments from all masked entries assessed, comprising the EoL value, represents the efficacy of that particular data set charting. Depending on the EoL value obtained for a nation-state or entity for a particular (masked) data set, a user of the NFAM would then have a quantitative measure for how accurate the data set will likely be when used to assess identities of unattributed entries. Further, the higher the X value and the lower the Y value were for a Hi/Lo Ratio, the greater the expected efficacy of the EoL and, therefore, the higher the confidence that the assessed identity of an unattributed entity was correct. Hi/Lo Ratios and EoL values calculated for each trail in this experiment are captured in Figure 2. Non-Forensic Attribution Methodology (NFAM) Trial Data Categories Leveraged in Each Trial and Resultant Hi/Lo Ratios and EoL Values (directly below).

One of the main reasons for the development of derivative EoL values, from Hi/Lo Ratios, was to easily (yet validly and reliably) support IC analysts' confidence in assessments made and to align NFAM quantitative outputs with values of the ODNI's Estimates of Likelihood (EoL) Matrix. If successful, such an alignment would allow those analysts less involved with data science and data processing to have a choice in providing either a quantitative measure of likelihood in assessments or a qualitative measure to customers who prefer these descriptors to numerical values. A full breakdown of estimative likelihood language and commensurate numerical value ranges within the ODNI's EoL may be found in Figure 2. The Office of the Director of National Intelligence (ODNI)-mandated Estimates of Likelihood (EoL) matrix (directly below).

Almost No Chance	Very Unlikely	Unlikely	Roughly Even Chance	Likely	Very Likely	Almost Certainly
Remote	Highly Unlikely	Improbable (Improbably)	Roughly Even Odds	Probable (Probably)	Highly Probable	Nearly Certain
01-05%	05-20%	20-45%	45-55%	55-80%	80-95%	95-99%

Figure 2. The Office of the Director of National Intelligence (ODNI)-mandated Estimates of Likelihood (EoL) Matrix
(*Note: Expressions of likelihood and probability must conform to ICD 203 analytic standards.)

Appendix B: The Cyber Lexicon for the Experiment

The following chart, Figure 4. The Cyber Lexicon for the Experiment (below), comprises the group of words defined particularly for describing activities found in the main deliverable and all included appendices.

Term	Definition (Suited to the Experiment)
Accuracy	The state of correctly identifying all (masked) malicious acts committed by a nation-state while limiting the number of misidentifications.
Actor (Malicious)	The proponent of the actions taken against the confidentiality, availability, and integrity of network operations, users, and data of the US government.
Category	One of four aspects of the (Cyber) Diamond Model (CDM); they are (1) Target; (2) TTPs; (3) Actor; and (4) Infrastructure.
Component	(1) Either of the two values (i.e. X or Y) in a Hi/Lo Ratio; or (2) synonymous with CDM Category.
Cyber Threat Framework	ODNI's mandated lexicon for describing and characterizing malicious cyber activity against US information networks.
Data	Fixed-phrase observations recorded in the DANTE database.
Efficacy	The level of accuracy a Hi/Lo Ratio reaches as to apply that result to a data set of unattributed entries for assessment.
Entry/ Submitted Entries	A row of data in the DANTE database relating the accounts of a malicious event against a US government information network.
Experiment	A test of activities, practices or procedures to determine whether they influence an outcome or dependent variable.
Explanatory (Research Design)	Model used to discover to what extent two variables (or more) co-vary, or how much changes in one variable are reflected in changes in the other(s)
Forensic (Confidence Pyramid)	A definitive assessment, highlighting the scientifically-based identification of the actor who committed a malicious cyber operation.
Infrastructure	Hardware, software, or processes leveraged by a malicious actor to conduct an operation.
Masked (Data)	Data (particularly entries) which have had the malicious actor who committed the transgression temporarily hidden for purposes of limiting bias during trials.
Methodology	A system (or collection of systems) used in the study of non-forensic attribution.
Outcomes	Quantitative or qualitative results from trials or calculations in this experiment.
Political (Confidence Pyramid)	A cursory assessment, highlighting the identity of a target as evidence of a malicious cyber operation.
Population	A group of individuals who comprise the same characteristics; in this case, nation-states.
Predictive (Prediction Research Design)	Model used to identify variables that will predict an outcome or criterion
Quasi-government Entity	US-based entities associated with the US government but which have authorities superseding US Title compliance.
Reliability	When individual scores from an instrument/tool are the same or stable upon repeated administration of the instrument/tool.
Sample/Sampling Frame	A subgroup of the target population a researcher plans to study to make generalizations about the target population.
Stage (reached) (in the CTF)	The level of success an actor had against a US government information network; the four stages are (1) Preparation; (2) Engagement; (3) Presence; and (4) Effect/Consequence.
Target	A victimized US Government information network, authorized user, or datum (data)
Technical (Confidence Pyramid)	A focused assessment, highlighting the operational aspects of targeted networks as evidence of a malicious cyber operation.
Trial	A test of the suitability of (a group of) variables to reach actionable levels of efficacy in attribution.
TTPs (Tactics, Techniques, and Procedures)	The manner(s) in which a malicious cyber actor prefers to operate.
Validity	Sound outcomes observed from the non-forensic methodology (or instrument used) which match or align with the stated purpose of this experiment.
Value(s)	Fixed scores assigned to (Cyber) Diamond Model-related data points
Variable(s)	A characteristic of something being observed which is both measureable and varies amongst groups or organizations.

Figure 4. The Cyber Lexicon for the Experiment

Appendix C: The Office of the Director of National Intelligence's (ODNI) Cyber Threat Framework (CTF)

The CTF is the cyber lexicon mandated by ODNI for use by Intelligence Community (IC) members (and potentially by all US government intelligence customers) to describe cyber-based activity, particularly malicious cyber activity against USG information networks. The lexicon was created and has since been managed by the IC's National Intelligence Manager for Cyber Issues (NIM/Cyber) to ensure that both intelligence analysts and customers had a common semantic understanding of terms used to produce deliverables within the cyber sphere regardless of each stakeholder's respective mission set.

The CTF consists of four stages, distinguished by the level of compromise achieved against a victimized information network. Preparation occurs prior to any activity on a network. Engagement represents a malicious actor's steps taken to initially get on a network. Presence is defined as having access to a network. Lastly, Effect is characterized by a disruption or denial of service to a network, data, or user. The four stages are also broken down by layer, of which there are three. Each layer down (from 1 to 3) represents a level of greater granularity in describing the activity observed, with Layer 1 being the most general and Layer 3 being the most technical. A chart of the CTF stages and layers is displayed below in Figure 5. The ODNI's Cyber Threat Framework (Truncated).

LAYER 1 - STAGES				
PREPARATION	ENGAGEMENT	PRESENCE		EFFECT
Actions to prepare to conduct cyber activities	*Actions to gain unauthorized access*	*Actions to maintain unauthorized access*		*Outcomes of actions on targeted system*
LAYER 2 - STAGES				
Plan Activity	Deploy Capability	Establish Initial Control	Establish Persistence	Deny access
Research & Analysis	Interact with Target	Hide	Expand Presence	Alter System Behavior
Resource/Capability Development	Drive-by attacks	Refine Targeting		Extract data
Conduct Reconnaissance	Exploit Vulnerabilities			Destroy HW/SW/Data
Stage Capabilities	Deliver Payload			Enable Other Operations
Initiate Operations	Suspicious Network Activity			
LAYER 3 - STAGES				
Review Strategy	Deploy Electronically	Unauthorized access	Anti-intrusion Detection Measures	Disrupt/Degrade Links
Plan Mission	Physical Proximity	Automated Malware C2	Anti-forensic measures	Disrupt/Degrade Network
Issue Guidance	Credential Farming	Establish Communications	Monitor Administrators	DDoS
Gather Intelligence	Social Engineering	Network Mapping	Increase User Privileges	Install Ransomware
Identify Targets	SQL Injection	Software Packing	Lateral Movement	Create Botnet(s)
Develop Infrastructure	Masquerade	Masquerading	Identify Targets of Opportunity	Deface Websites
Physical Reconnaissance	Use of Exploit Kit	Obfuscate Payloads	Service Manipulation	Relocate/Store Data
Electronic Reconnaissance	Cross Site Scripting	Indicator Blocking	Registry Run Keys	Disclose Data
Stage Externally/Internally	Webmail Vulnerability	Scripting	BIOS Rootkit	Exfiltrate Data
Issue Operational Tasking	App Vulnerability	Disabling Security Tools	Master Boot Record	Establish C2 or Hop Point

Figure 5. The ODNI's Cyber Threat Framework (Truncated)

[i] John W. Creswell; Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th ed.); 2012.
[ii] John W. Creswell; Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th ed.); 2012; pg 337-342.
[iii] DHS; Internal Database; accessed on 18 May 2017; DOI: 2015 and 2016; Extracted information is UNCLASSIFIED; Overall database classification is Higher in Classification than Document Classification..
[iv] John W. Creswell; Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th ed.); 2012.
[v] John W. Creswell; Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th ed.); 2012; pg 337-342.
[vi] John W. Creswell; Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research (4th ed.); 2012; pg 157.

Other References

This document is provided for educational and informational purposes only. The views and opinions expressed in this document do not necessarily state or reflect those of the U.S. Government or the Exchange Program Partners, and they may not be used for advertising or product endorsement purposes. All judgments and assessments are solely based on unclassified sources and are the product of joint public and USG efforts.