Efficient Reactive Forensic Investigation Model for Large APT Computer Network Events

***Abstract*—Forensic investigation is an evidential process that occurs after a computer network security event to ascertain causes, vulnerabilities, and remediation actions. The security problem is establishing evidential providence when Advanced persistent threats (APT) are deceptive, learn new behaviors, and erase evidence of activity. Proactive defense mechanisms such as honey pots, fake files and so on filter many APT attacks but network security also relies on reactive forensic investigation after security events to learn tactics and to harden the system. Root cause analysis of APT activity is frustrated by gaps in the data, similarity of normal and malicious network processes, confusion with other network attacks, and the evasive behavior of an APT. In this research the problem of APT evidential providence in reactive APT investigation is addressed by innovating an APT network forensic investigation model (APT-FIM) solution to differentiate reactive APT evidence from other evidence in a computer network investigation. The APT-FIM is structured from the APT definition using Object Oriented (OO) methods and designed for practical use by a security practitioner on secondary data sources. The contribution of the research is to formulate efficient methods and effective big data guidance for reactive APT investigation. Two use cases are used to assess the model and to identify missing requirements. A practitioner flow chart for use results.**

***Index Terms*—Advanced persistent threat, Forensic investigation, Network security, Evidential providence, Big data evidential efficiency, Practical advice**

I. INTRODUCTION

THIS research was motivated by the need to improve data processing accuracies and efficiencies in forensic investigations for large commercial service networks, where large data volumes, and large quantities of network processes activity require effective investigation data management methods. The distinction between proactive security defenses and reactive investigation security defenses is made to target security compromise within computer networks and innovate a solution for the threat. Security compromise is active in all network planes and the use of deception for access, the greatest current vulnerability for computer network security. The advanced persistent threat (APT) is a well-documented strategic threat to network security that uses semantic deception as one of the attack and tactical vectors [1][2]. Evidential providence is the challenge and problem for investigation when an APT has an abstract high-level definition (Tab. 1 from [4]) but ambiguity for identification in computer networks. Other attack definitions, usual network services, obfuscation actions, and deception ploys all use the same computer network services. The APT also harnesses many of the mechanisms of other attacks, the camouflage of big data, and the usual functional services of a computer network to achieve the mission aims [3]. Hence, the APT investigation challenge is to differentiate the sub-class of features that identify an APT attack and the classes of features that are also common to other network activity. The foundational definition of an APT [4] describes a potential threat to networks that is characterized by staged management from a skilled network expert knowledge database using advanced combinatorics or sequences for effect. The second APT characteristic is that it is persistent in that it may continue over extended periods of weeks, months, and years, to achieve the adversary's mission or purpose. The third APT evidential characteristic is that it is a threat to high value information and has a high probability of mission success (Tab. 1 from [4]). The definition is helpful to filter out script-kiddies, one off attacks, and low-value targets but remains a high-level abstraction that may only be known by the effects of network attack. Therefore, our research contribution is to model the consistencies specified in the definition for root cause analysis, security event investigation, and evidential providence. The APT-FIM model enhances reactive network investigation by differentiating APT activities from other attacks and normal network operating evidence.

Our APT-FIM model (Fig. 1) uses the principles of Object Orientated (OO) design to map the features of the APT formal definition, the features of computer network management, and the abstraction that has all network security features. These features are grouped for similar objects into five class layers and the related degrees of certainty for APT identification decision-making. When applied the model has three decision outcomes (1) certain APT feature match, (2) similar APT feature match, and (3) uncertain APT feature match. The OO design has the upper layer of all security abstractions, and a sub-class for APT abstractions to group the attributes of an APT and create logical grounds for an inheritance mechanism to determine the match of network evidence to the APT definition. Hence, the upper layer is deterministic for causal chain links, and proof of evidential providence. The second layer is polymorphic and where semantic meanings are determined. Hence, it includes a probabilistic element to weigh the degree of uncertainty a concept or linguistic term possess to the APT definition by considering potential instances of multiple

a

inheritance within the broader class of all security abstractions. The third layer groups all computer network process objects which may be used by an APT or any other attack or normal network traffic. The fourth layer groups all network activity objects which may be used by an APT or any other attack or normal network traffic. The fifth layer groups all network metric objects that report the physical and logical performances of a network, such as time stamps and traffic quantities [5]. Each layer of the APT-FIM model (Fig. 1) provides levels of confidence to help an investigator determine the relationship of the evidence to an APT attack.

Layer		Decision Failure Risk
1	All Security APT Abstractions Abstractions	Low
2	Semantics	Medium
3	Processes	High
4	Activities	High
5	Metrics	Medium

Fig. 1. APT-FIM model for APT evidence determination.

The model fits zero trust environments in which semantic deception is active [6]. Semantic deception takes many forms with one of the predominant methods being exploitation of authorized end users or similar code to gain network access. The deception then exploits normal network work systems to action and deliver the mission of the attack. Proactive defenses may stop some of these attacks, but other attacks get into the network and cause security events [7]. The investigator is left with effects of the attack, for example the use of exfiltrated information to steal money from a bank account, but little else when the attacker erases evidence and conceals the evidential trail in legitimate processes and work system [8]. The APT evidential trail within a network is entwined with many other competing processes making it incomplete and factually challenging to accurately establish.

A computer network investigator has the network architecture, the appliances, the devices, and the applications in which to search transaction logs and data repositories [9]. In large commercial networks these are big data conditions, and we propose the APT-FIM model to increase efficiency by using better logical techniques for time cost efficiencies and increasing APT identification accuracy rates. At present, an investigator is challenged to distinguish the true from the false when investigating network APT security breaches and is prone to miss evidence, misinterpret data, confuse legitimate processes with malicious use of the same process, and waste resources on false positives. The APT can be a perfect attack that exfiltrates the required information from a network but then leaves without trace [1]. The following sections report previous APT research, the APT-FIM model design methods, the application of the model to two use cases, and the findings. The use case analysis identifies four further requirements. These requirements are formulated and a practitioner guideline for use completed.

2. Background Research

Publications on APT network security research are categorized into proactive and reactive reports. Proactive reports focus on strategies aimed at preventing attacks through deception and intelligence, while reactive reports have research on the effects of APTs after a network attack [11]. Proactive research seeks to thwart APTs from infiltrating a network using threat intelligence and deceptive semantic methods, which may involve tactics such as generating dummy files, anonymization, and deploying honeypots. In contrast, reactive research acknowledges that APTs can bypass even the most robust security measures and concentrates on activities such as detection, monitoring network traffic and files use patterns, and predicting APT lifecycle behaviors. The literature review employed keyword searches to filter and analyze 50 relevant papers published between 2014 and 2023, which discussed APT research from five IEEE and SCIMANGO Q1 & Q2 ranked security journals. The following subsections offer the foundational definition of an APT, conduct a definition analysis, and review current APT research publications that encompass both proactive and reactive APT security preparations.

2.1. Advanced Persistent Threat Definition

Reference [4] defined the APT threat and [10] has clarified the APT risk context and general parameters for the threat to computer network systems. They describe an APT as:

> "An adversary with sophisticated levels of expertise and significant resources, allowing it through the use of multiple different attack vectors (e.g., cyber, physical, and deception) to generate opportunities to achieve its objectives, which are typically to establish and extend footholds within the information technology infrastructure of organizations for purposes of continually exfiltrating information and/or to undermine or impede critical aspects of a mission, program, or organization, or place itself in a position to do so in the future; moreover, the advanced persistent threat pursues its objectives repeatedly over an extended period of time, adapting to a defender's efforts to resist it, and with determination to maintain the level of interaction needed to execute its objectives."

Analysis of the APT definition [4] in Tab.1, clarifies what the abstract APT definition means in practice for security practitioners by mapping the APT definition characteristics to identifiable threat features.

TABLE 1
APT Characteristics and Features

APT Characteristic	Identifying Features		
	Abstraction (Referrals)	**Semantic (Methods)**	**Attribute (Indicators)**
1. An adversary with sophisticated levels of expertise	Competitive Warfare [4]	Expert Intelligence [28]	Reputation [31]

2.An adversary with significant resources	Enduring Capability [30]	Strength [29]	Numerical advantages [21]
3.Multiple different attack vectors (e.g., cyber, physical, and deception)	Complex Identity [7]	Multiplicity of co-ordinated attacks [8]	Multi-factor impact [19]
4.Generate opportunities to achieve its objectives	Innovates Success [18]	Creative Methods [44]	Vulnerability exploitation [39]
5.Establish and extend footholds within the information technology infrastructure	Tactical Deployment [16]	Permanent presence [30]	Countable Occurrences [5]
6. Continually exfiltrating information	Perpetual Action [8]	Active agency [21]	Services use [35]
7.Undermine or impede critical aspects of a mission, program, or organization	Targeted Disruption [42]	Goal & objective compromise [46]	Successful failures [4]
8.Position for future action	Strategic Deployment [40]	Action ready [30]	Concealment [41]
9.Pursues its objectives repeatedly over an extended period	Timeless risk [11]	Certain objective success [14]	Event count [33]
10.Adapting to a defender's efforts to resist it	Evasive learning [18]	Avoids defences [23]	State changes [44]
11.Determination to maintain the level of interaction needed to execute its objectives.	Active state agent [9]	Intentional network use [35]	Transaction rates [45]

2.2. Proactive Threat Intelligence APT Methods

Traditional network security theory divides a network into three regions of a green internal zone of trusted network use, a firewall and authorization system barrier to filter all external traffic, and a red zone for all untrusted traffic [12]. Proactive threat intelligence focuses on the red security zone to analyze and learn potential attack profiles and to inform strategy for network protection. The literature reviewed identifies four methods for APT threat intelligence gathering: Games, Deception; Modeling, and threat Actor profiling [Fig.2]. Each method allows disclosure of intentional information and providence for traffic entering a network. The methods also delay and disrupt a potential attack by withholding access rights or satisfying the attacker with worthless information.

Reference [13] advocates dynamic Game theoretic strategies as a proactive defense against APTs. The definition of an APT describes evasive and intelligent behaviors that are capable to evade and to learn countermeasures for static and standardized security measures, hence games are structured to entertain and trick all traffic into game play before access rights might be authorized. Traffic that exhibits APT defined behavior may be played in the red zone until exhaustion and intelligence gathered to inform stronger network protection. Reference [14] suggests obfuscation of adjustments to network security improvements by engaging attackers in intelligent games that cover any adjustments to posture and mechanisms so that the attacker is caught out by making incorrect player moves. Reference [15] defines two evolutionary games to trap attackers into a state of equilibrium where the attack intentions are frustrated, and they leave to seek other targets. Reference [16] elaborates game strategy for APT defense that raises the cost of attack while offering more satisfying information exit points from the game. These strategies are proactive and engage the APT attacker in simulated success scenarios that satisfy an intelligent attacker who will rationalize costs of continuing against the benefits of taking the information prizes offered, while unintentionally disclosing identity and behavioral information.

Reference [17] advocates the use of a dynamic deception framework to prevent APTs entering a network. The framework incorporates IP hopping, Honey pot diversion, and data encryption to confuse an attacker and to frustrate APT attacks. Reference [18] takes a different approach and studies the evasive maneuvers an APT can make to avoid detection and proposes algorithms that know predictable APT behaviors for structured deception phases to deceive an attacker. Reference [19] proposes a disinformation model for APT counterattack. The model is targeted at threat actors or third parties that benefit from the activities of an APT by providing convincing and similar disinformation for exfiltration. Reference [20] advances network defense strategy by providing Honeypots of interesting information to satisfy or deceive the attacker.

Many different models are proposed to represent the different phases of an APT attack [21][22][23]. Proactive intelligence is collected from an APT in the first phase of APT attack when the APT places a network under external surveillance for intelligence gathering. Network security requires situational awareness for these activities and to collect information on all probes and fake digital personas interacting with the network [24]. Matching of patterns and cross-referencing to IP, Hash, and alert libraries, assists with profiling and identification of threats, and the classification for defensive strategy development. Reference [25] advances Artificial Intelligence and statistical models to enhance network security awareness in the red zone. The models use learning and statistical correlation methods to predict the presence and attack vectors of an APT before the opportunity to impact the network. Reference [26] takes a comprehensive approach to threat intelligence modeling and uses the information gathering phase for attacker vulnerability and disclosure of valuable counterintelligence.

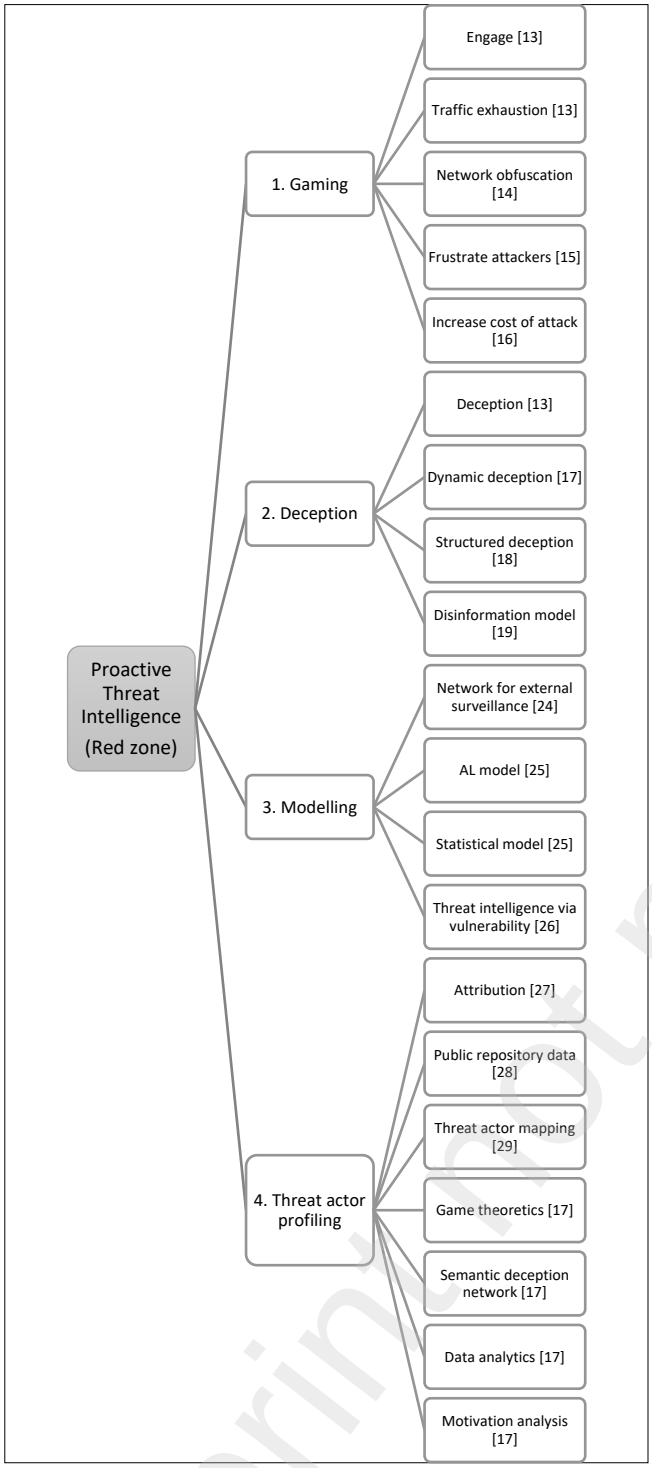

Fig. 2 Proactive threat intelligence methods

A substantial body of research concerns identifying and locating the attackers and the beneficiaries of an APT attack. There are both proactive and reactive research reports. The reactive methods concern attribution of responsibility for attack from threat intelligence [27], and data analytics on publicly available databases of archived information [28]. Reference [29] defines a taxonomy of persistent techniques that are mapped to specific threat actors. The identification of a threat actor opens courses of action beyond the network and to legal and economic remediation. Disclosure of threat actors comes through game theoretic tactics, semantic deception networks, data analytics, and motivation analysis [17]. Threat actors employ semantic methods as the means to host APT attacks, with the endgame to achieve material and reputational goals [30]. The identification of the actors as layers of mutually dependent attackers and beneficiaries broadens the scope of network event investigation and differentiates proactive intelligence gathering methods from the reactive.

2.3. Reactive and Post Event APT Methods

The research publications reviewed predominantly concerned tools and techniques for reactive detection of APTs. The success of APTs to penetrate network defenses was undisputed and the requirement to respond to network security events standard security practice [31]. Reactive methods are required to restore the network state and have the benefit of learning from past security failures. The predominance of reactive research also suggests that APTs are difficult to predict and become more accessible for study from network artefacts and forensic evidence than prior intelligence [8][32]. The detection and discovery of APT activity in the network green zone is a common research theme in the literature reviewed. Three methods are advocated: reactive, defensive, and investigative [Fig.3]. The following paragraphs summarize research contributing to each method.

Seventy-two percent of the research papers reviewed reported reactive research into APT activity for green zone network security improvement. The research topics were artificial intelligence and deep learning for APT detection [33], traffic volume and exception analysis [34], behavior monitoring [35], big data analysis [36], and APT lifecycle discovery processes [37]. Each report selected one or some of the APT characteristics listed in Tab.1 but failed to investigate others. It was also often not clear how the decision to discriminate APT traffic from other traffic in the network was determined. Most reports used pre-classified data that was researcher generated or downloaded from libraries. Reference [38] specifically cautions limitations for APT research based on these data sources. The failing is that the advocated detection and classification methodologies may underdeliver in different conditions and not give the intended security protections. These limitations and APT definition coverage gaps in the literature open possibilities for further research and expanding the APT reactive research agenda to be inclusive of all the APT characteristics in Tab.1.

The reactive research reports reviewed [Fig.3] advance methods for better defending networks from APT attack [39][40]. This includes tactical requirements for network defense. Reference [41] advocated analysis of network events to identify root causes and the subsequent network remediations. Reference [42] promotes methods for causal analysis of APT attacks to improve defensive awareness and network readiness to resist attack. Discovery methods are also used to inform defensive strategy and tactics [43]. Traffic analysis research employs exception analysis to alert APT activity [44], and [45] employs big data methods to identify and propose security remediation for APT attacks. Reference [46] focuses on APT behavior to inform defensive mechanisms that alert a network security function to disrupt an activity and store the behavioral patterns for future reference. The research

reports advance methods for network defense from APT attack but have gaps in APT characteristic coverage, management of uncertainties, and clarity of definition for semantic deception variations.

Incident and event investigation is a key task for network security improvement. Reference [47] argues that three domains of knowledge converge to make an effective APT investigation strategy. First, the APT ontology, second the general security domain knowledge, and third the network domain specific knowledge. These three domains form a framework in which evidence is triaged for acceptance or rejection in an investigation. The target objects in APT investigation are both static for stored data and live for traffic flows. Reference [48] scopes the challenge, [50] frames the context for APT investigation, and [39] reports recovery and remediation actions.

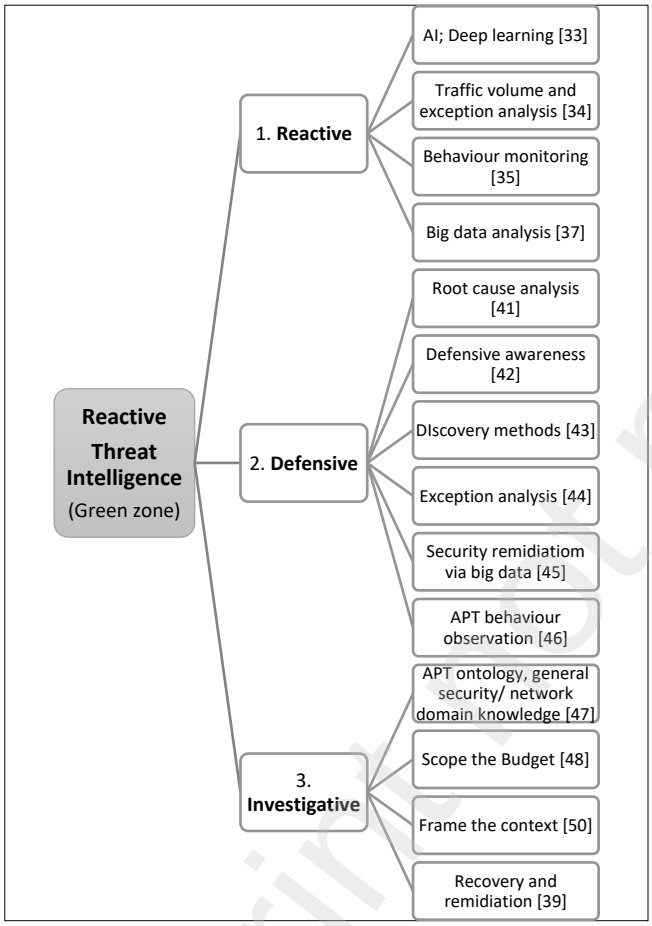

Fig. 3 Reactive threat intelligence methods

3. Reactive APT-FIM Model Design Methods

Object Oriented modelling in software engineering follows the logical steps of abstract descriptions of the problem, solution designs, code, testing and implementation [51][52]. This research reports the problem contextual abstractions, the model design, and in Section 4., two use case examples. The coding, final testing, and implementation are out of scope for this conceptual research report. In Section 3.1. data structures are defined for the theoretical and materiel objects in relation to physical computer networking, computer network security, and APT risks. Also, a data ontology is specified so that evidential providence may be established by inheritance. The classes are selected from the context for general similarities, and sub classes for specific similarities. Section 3.2. defines the scope of the APT-FIM model to be for reactive investigation data (Section 2.3.) and excludes proactive investigation data (Section 2.2.). The data methods and attributes consist of the APT reactive definitional elements and APT behaviors. Hence, the conceptual model design architecture is completed. Section 3.3 specifies the semantic discrimination mechanism that distinguishes APT evidence from other network attack data and normal network use data. The model has three decision outcomes for evidence triage, (1) certain APT feature match, (2) similar APT feature match, and (3) uncertain APT feature match. The theoretical accuracy and risk tolerance bands for the model are calculated in Section 5.2

3.1. APT-FIM Structures

Reference [47] defines the APT investigation context as having three domains of dependent knowledge that objectify the targets for investigation. The objects of investigation are both theoretical and materiel and have distinct methods and attributes for evidence collection. In this research the abstract theory domains define security and APT data, whereas the materiel assets of network store, process, and communicate data for examination. Instances in each domain of knowledge are data objects for grouping into general classes and specific sub classes. Hence, model classes are selected for the set of all security abstractions, and the security sub-class described by the APT definition [Tab. 1]. The network context has four materiel elements of hardware, software, humans, and services [39], that scope an investigation. The four elements contain data objects that group into classes of network processes (logical), network activities (physical), and network metrics (logical). The APT-FIM model is a problem solution hence, these classifications provide structure and hierarchical relationships for evidential providence. Each process has many associated activities, and each activity has many metrics to measure network performance. Fig.4 summarizes the APT-FIM problem condition and the classification of probable object classes.

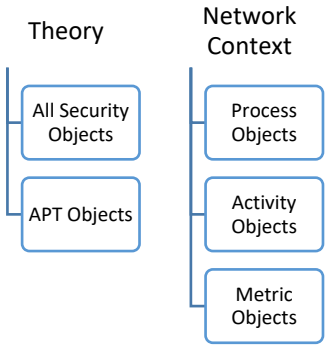

Fig. 4 Object Class Selections

APT evidential providence is established by relationships between the classes. APT objects are a sub-class of all security objects and inherit definition and meaning from the parent class. However, the gap between theory and the materiel network context is polymorphic and introduces elements of uncertainty. Hence, there are many more security objects that inherit and use the Network Context objects and the related methods. Therefore, we introduce a mediation layer termed the Semantic Layer to discriminate APT and non-APT evidence by

meanings derived from the homomorphic objects in the APT definition [Tab. 1]. The Semantic Layer is required to select evidence that satisfies the APT theoretical definition and to exclude evidence that has a polymorphic condition. The risk is the exclusion of lower-level evidence when for example, the APT manages and uses other security objects for tactical advantage, and normal network traffic is also using the same methods. However, the problem is an economic trade-off, and a network investigation must be cost effective. Degrees of uncertainty are managed in the model with the assigning of probabilities in relation to providence, and cost efficiency by choosing the highest probabilities.

Fig.1 shows the APT-FIM as a layer hierarchical model with the Semantic Layer mediating between the theoretical objects and the materiel network objects for evidential trust management. The Process and Activity Object Layers are generally untrustworthy for definitive APT evidence, but the Metric Object Layer has Objects that are useful and sufficiently stable for APT investigation. Network clocks and time stamps for example, are global objects that are difficult to corrupt and have metrics for timeline creation and systematic metrics to improve investigation efficiency. Whereas Metric Objects for traffic are less trustworthy but may be used as indicators or triggers for investigation. Fig.1 also provides a risk estimation scale for APT evidence discrimination failure. The abstract definition of APT [Tab.1] provides certainty for evidential providence, but Semantics and Metrics fall into a grey area where some methods and attributes satisfy the APT definition, but others do not. Hence, the application of the APT-FIM model in network investigation requires further features to assign degrees of uncertainty, the mechanisms for semantic discrimination, and triage categories. These requirements are addressed in the following sub-sections.

3.2. APT-FIM Data Methods & Attributes

The scope of this research is reactive forensic investigation and excludes the classification of proactive methods and attributes. It also has the constraints of green zone and firewall boundaries. Reactive APT investigation in large commercial networks is challenged by the big data issues of data volume, data velocity, and data veracity. Hence, efficient investigation methodology has economic and trust controls that exclude budget overruns and evidence sources that are untrustworthy or have a low probability of satisfying evidential providence criteria. The objective of APT network traffic investigation is to classify, organize, and store the data that fits the APT descriptive case and to reject other representations. In the APT-FIM the theoretical class of APT Objects is trustworthy whereas the Network Context Classes are polymorphic, and evidence is excluded or given a low priority probability. In this way a network APT security event investigation confidently delivers trustworthy evidence from a big data context without wasting resources on less valuable lines of inquiry. The exception is the Network Context Metric Class that has clock and time methods and attributes, that are globally consistent across a network.

The inclusion and exclusion of classes from Fig.4 into a network security event investigation reduces the number of indeterminant evidence sources and increases those with a high probability of APT providence. Hence, APT Object methods and attributes for reactive investigation describe intentional homomorphic behaviors consistent with the APT definition rather than attribution to generic processes, activities, and metrics that are used by all attacks and all normal traffic. Tab. 2 summarizes these reactive APT features [Tab. 1 items 8-11] that differentiate inherited patterns of behavior from other behaviors in network system.

TABLE 2
APT INHERITED REACTIVE METHODS & ATTRIBUTES

APT Object Methods	**APT Object Attributes**
innovatesSuccess()	Exploitation
tacticalDeployment()	Countable instances
exfiltrateInformation()	File & services for export
targetDisruption()	Successful failures
strategicDeployment()	Concealment
perpetualActions()	Incident rate count
evasiveLearning()	State changes
activeAgency()	Transaction rates

3.3. APT-FIM Semantic Discrimination Mechanism

The Network Objects in Fig.4 give little confidence that any process, activity, or metric may be attributed easily to an APT attack in causal analysis when an APT uses the same network processes and activities as other attacks and normal traffic. In large commercial networks the predominant volume of network traffic comes from normal workloads and to a lesser extent wide-ranging attacks and illegal network use. In these conditions determination of APT evidence must have an effective discrimination mechanism. The predominant mechanism in forensic investigation is triage where evidence is classified into three categories [Fig. 5] based on relevance to the investigation objectives. However, in the case of APT investigation further resolution is required to distinguish from similarities, fakes, and deception. The APT-FIM model uses homomorphic and non-homomorphic object differentiation to resolve meanings for theoretical security Objects in relation to network context Objects. The Semantic Layer inherits from All Security Objects and the sub-Class of APT Objects to discriminate meanings with a mechanism for semantic determination.

The APT-FIM hierarchy provides evidential providence by inheritance and the opportunity to discriminate objects as evidential, non-evidential, and indeterminate on scales of trust probabilities. For example, malware is a security Object that has theoretical definition and may be discovered generally in the network context. The sub-Class APT objects may use malware for any of the eight reactive Object methods [Tab.2] to achieve mission. The challenge for the semantic discrimination mechanism is to determine the status of the malware evidence in relation to the APT definition. In the case of malware, it can be unrelated to APTs, managed by APTs, or judged as an indeterminant probability in relation to APT Objects. Investigation triage then has a three-way trilemma that is satisfied by a systematic resolution solution [Fig.5].

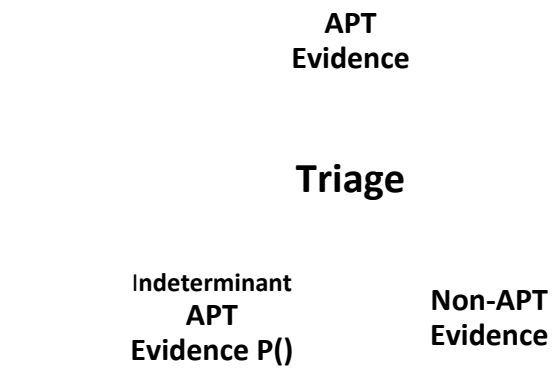

Fig. 5 General APT Triage Discrimination Mechanism

The first resolution step has eight affirmative questions derived from Tab.2 to triage reactive investigation evidence as APT:

- Are creative methods found?
- Does the evidence show a permanent network presence?
- Were the channels continually active?
- Have system goals & objective targets failed?
- Did strategic advantages materialize?
- Is there a timeline for the successful attack?
- Is there evidence of avoidance behaviors?
- Is there evidence of intentional persistent behavior?

In investigation APT evidence may not be sufficient for certain and affirmative answers, and then the remaining evidence requires resolution in Step 2. Step 2 assigns trust probabilities to any evidence that is not accepted as APT in Step 1 by assessing similarity to the APT definition on a scale of 0 to n, where n is $0<n<1$. A further constraint on accepting indeterminant APT evidence is the economic cost of managing large data sets. Hence, the investigation budget sets a scale for probability value cut-off, and the remaining data is triaged as non-APT. To assign probabilities the semantic mechanism estimates the closeness of the evidence to the sub-Class APT data Objects. For example, a Metric Object timeline and time stamps would be determined APT evidence in Step 1, but traffic variation evidence rejected as variations are by association and with published margins of error. Hence, Step 2 assigns the probability that the evidence is related to an APT attack and checks the probability value fits the budget scale range. The association of APT activity with network Traffic metrics is ranked based on the reliability and validity measures of the metric tool or discovery method used to measure. The investigation appetite for risk determines which evidence can be used in the event Report and which evidence is rejected based on the estimated trust probabilities. Hence, in two steps all evidence is triaged. The semantic mechanism determines the value of all evidence in relation to the APT theoretical definition, and investigation objectives.

4. Use Case Application of APT-FIM

A use case describes a function that a system model performs to achieve the user's goal. It must deliver an observable result that is valuable to potential model users. Two APT scenarios are selected from live computer network forensic Reports in two large commercial multi-national networks to test the model performance and to identify further requirements. The identities and data from the commercial enterprises are anonymized and hidden for non-disclosure. The APT-FIM is applied to reactive secondary evidence to decide if an APT event has occurred. In both cases the attack vector of semantic deception was used to breach network security provisions, but it is uncertain if this is an APT or unrelated attack. The value of obtaining a determination from the APT-FIM is that the correct countermeasures can be applied to protect the computer network. Often an APT attack remains undetected and is a latent unpredictable security event waiting to happen. The first scenario is from a Banking system and the second a Government department.

4.1. Banking System APT Attack Use Case

The bank network transacted 700 to 1000 transactions per second (TPS) making it a large commercial network. Customer services notified the network security Unit that there was a complaint that money had been transacted in an account from a remote location without customer authorization. The network security Unit initiated incident response procedures and searched the bank network and systems for evidence. Other transaction anomalies were identified in dormant accounts, and the residual code used for semantic deception of users located. The incident was escalated to a security event. The problem faced by the network security Unit was to determine the scope of intelligence behind the event and if all the isolated security violations formed a coherent pattern with causes to explain the documented effects. External security consultants were called for help, and the APT-FIM used to triage the secondary data and determined evidence in scope for an APT attack.

Table IV is the use case analysis and demonstrates the application of the APT-FIM on the secondary data found in the network security Unit event investigation Report. The security management problem was to decide if the well-orchestrated incident chain from a customer victim response to semantic deception code, to dormant account balance alteration, and to the use of ATM card transactions from a remote hard currency location, fitted a bigger pattern for network security compromise and signaled potential further green zone data and reputation losses. The escalation to a security event could be justified by the identified 336 dormant account compromises in 6 hours, but the full scope of the event remained unclear and fragmented into isolated clusters of data. The network security Unit needed to know if other attacks were imminent or if the capability for further security compromises was already in the network green zone. The APT-FIM was applied to the network security Unit event investigation Report using the eight reactive APT definitional metrics. The Report held copious data on network architecture, log files, transaction records, date & time stamps, and so on. The eight high-level APT-FIM metrics are a tool to speed decision-making and then if further evidence is

required probabilistic methods and full budget estimates apply. Questions arose as to how many repetitions of the same APT attacker behavior were necessary to flag an APT event, or how many metrics had to be satisfied for a definitive decision. This information is absent from the model and further specifications are required in the APT-FIM for user guidance.

In the absence of clear guidance users persist and identify as many as possible APT behaviors in the data sets, introducing wastage and gut estimate risk. The size and scope of this commercial network had led the investigators into the big data trap of being comprehensive when a decision to flag a security event of a strategic scale could be quick and easy to call from a minimal data set. All metrics were quickly affirmative and one or two repetitions sufficient from standardized data sets. Learning from the APT-FIM use case provides recommendations for improvement and practitioner guidance. The findings of Tab.3 require urgent but standardized APT security remediation actions to deeply clean the computer network, and to review general security policies.

TABLE 3
USE CASE 1 APT-FIM METRICS

APT-FIM Metric	Y/N	Evidence
Does the evidence show a permanent network presence?	Yes	The victim executed semantic deception code 5 months before incident.
Were the channels continually active?	Yes	Continuous SSH reverse tunnel exfiltration to by-pass firewall.
Have goals & objective targets failed?	Yes	Bank reputation, information, and financial target losses.
Did strategic advantages materialize?	Yes	Attacker financial and reputation gains. Bank public reputation loss.
Is there a timeline for the successful attack?	Yes	Five months: Semantic deception perimeter breech. Exfiltration. Escalation of privileges. Attack.
Is there evidence of avoidance behaviors?	Yes	Gaps in logical sequences of log files suggests erasure.
Is there evidence of intentional persistent behavior?	Yes	Multiple server passwords scanned from within green zone over 3 months. Cts. exfiltration of information.

4.2. Public Agency Use Case APT Attack

The public government agency served a population of general citizens and networked with the federal government information and transaction services system making it a large geographically distributed service network. When employees logged onto their Public Services computer network a Ransomware demand screen appeared. The IT security Unit immediately escalated the attack to a security event as the whole network was impacted. The scale and impact on services required external consultants to investigate and to remediate the event. The consultants investigated and documented the Ransomware event, restored services, and provided detailed advice to defend future attacks, but the Public Agency wanted to know if this event was a one off or part of a larger orchestration of attacks on the computer network system. The APT-FIM was applied to the secondary consultant data to determine evidence in scope for an APT attack.

Tab.4 is the use case analysis and demonstrates the application of the APT-FIM on the secondary data found in the consultant's Ransomware event investigation Report. The security management problem the Client faced was to decide if Ransomware evidence and related causal factor analysis could be attributed to a more general and comprehensive attack on the computer network services. The remedial and security review guidance for the Ransomware attack addressed specific instances of security failure but lacked coherence with previous incident and event reports. The APT-FIM identifies evidence consistent with the APT definition in a computer network green zone or firewall by ranked triage of data. Primary or secondary evidence is triaged to determine coherence with the APT definition, and a positive determination signals hidden network vulnerabilities for potential future security events. The application of the APT-FIM to the Ransomware investigation Report was fast and efficient. Step 1 eight higher-level metrics were applied to the secondary data sets to identify specific behaviors and targeted APT artefacts. The process of triage required no more than 10 minutes because Report evidence satisfied a positive assertion to each metric, immediately signaling an APT event and no further action was needed to triage any further evidence. The same user guidance gaps of use case 1 were noted but also the lack of controls for information release required redress. The results are summarized in Tab.4 and standardized APT remediation and security policy review advised.

TABLE 4
USE CASE 2 APT-FIM METRICS

APT-FIM Metric	Y/N	Evidence
Does the evidence show a permanent network presence?	Yes	Executed .exe file in system for at least 6 months. Fake user 3 months old in use.
Were the channels continually active?	Yes	At least 18 months. Multiple channels.
Have goals & objective targets failed?	Yes	Work system shut down. Payment system unusable. Reputation damage.
Did strategic advantages materialize?	Yes	Information exfiltrated prior to ransomware. Calculated reputation loss.
Is there a timeline for the successful attack?	Yes	Multiple orchestrated incidents over 18 months.
Is there evidence of avoidance behaviors?	Yes	Anti-virus disabled or protective exclusions inserted to allow access.
Is there evidence of intentional persistent behavior?	Yes	Persistent remote login attempts. Multiple related attacks over 18 months.

5. Requirements Learned from Use Cases

The APT-FIM use cases identify further requirements for the model. Logically a computer network forensic investigation completes data collection before the data analysis phase. Hence, the APT-FIM is used in the data analysis phase for determination of a specific attack from reactive green zone or firewall secondary data. Low accuracy rates occur when attack effects cannot be traced to root causes or quantified attack definition metrics. In practice a gap exists between theoretical descriptions of threats and the empirical computer network attack(s) that satisfy the descriptions. The APT-FIM resolves the problem for APT attacks by modeling logical and physical context classes, but also creates a semantic class for objects to mediate the theory-practice divide. In the following sub sections the additional requirements for the use of the APT-FIM are defined, the computational risk requirements and theoretical limitations for the model specified, and the impact of investigation budget constraints for big data reviewed.

5.1. APT-FIM Additional Requirements

A forensic examination requires cost efficiency at each phase. In big data environments the constraint is critical for resource management and budget compliance. In Case 1, the Bank internal network security Unit had completed a thorough investigation of the security breach and discovered the semantic deception code, its execution, multiple anomalies that subsequently began to occur in the green zone, a backdoor to the red zone, a definitive timeline, and proactive evidence leading to actors. However, the event could be an isolated incident and unrelated to an APT attack. The challenge was to apply the APT-FIM to the use case event Report data and to conclude an outcome. However, no guidance was provided as to how many affirmative data constituted a yes, how many of the eight questions had to be yes, or when to stop processing data. These are important requirements if the model is to be cost effective in big data contexts. The investigator also requires guidance to limit the amount of data processed before drawing a conclusion.

The APT-FIM was efficient in the use cases because the vast amounts of evidence had been pre-structured into Reports by investigation aims, objectives, and data collection methods. The Reports were also authentic, comprehensive, and in standard industry format so the data could be trusted as representative of the computer network states. The rapid use case conclusions were gained by gut estimation of the amount of data to process for an affirmative response (P(X)=1) to each APT metric but no consideration was given to margins of error, tolerance boundaries, or how trust probabilities are assigned when P(X)<1. These are important requirements if the model is to be accurate and repeatable in use. When the Reports have variable formats, unusual methods, unexplained data omissions, and so on then each Report requires an audit mechanism before use. Another layer of analysis is required to triangulate findings with trusted data sources, and an alert signaled to APT-FIM users for secondary data verification before use.

To be effective the users must be able to trust the APT-FIM results. The design of the model creates providence by selecting homomorphic objects from the APT definition and then comparing all other objects. The hierarchical structure enhances progressive resolution of evidence against the homomorphic objects to quantify units of trust. In use case 1 the APT-FIM gained a rapid decision response time by data reduction so that the security Unit had certainty with regards to the APT nature of the event, and the trust to remediate and to apply APT countermeasures. However, in use case 2 trust in the model had to extend beyond the technical user community and into the public domain. The organization's reputation was seriously damaged through media scrunty of the ransomware attack and public loss of service confidence. The event Report further eroded public trust in the state government with disparate technical details that confused non-technical people. The issue of communication of outcomes from the APT-FIM to beyond the technical security community has not been considered. The public release of information that a security event is also an APT attack with a high likelihood of further disruption to services, requires user alert and management information disclosure processes. Hence, an additional requirement is to alert APT-FIM users to the risks of information disclosure.

The APT-FIM is designed for experts to use but it has the potential to be generalized and automated. Such improvements require the development of a simplified user interface and end user guidelines. Further conceptual modelling can enhance the range of potential questions the APT-FIM may answer. In the two use cases the model was useful because it gave a binary answer to the APT attack definition, yes or no. To further triage evidence and assign trust probabilities requires quantification for calculating degrees of uncertainty in relation to the homomorphic APT objects as elaborated in Section 5.2. Consistency for general use of the model shall require the classification of evidence type, risk scales for probability estimation, and further use case examples for user guidance. Automation is initially for supervised use where the investigator has sufficient learning from use case examples to apply live automated discovery processes. Fully autonomous implementation of the model is beyond the scope of current research, but the semantic discrimination mechanism has sufficient structure for machine learning algorithms and application to primary data. Generalization and automation are two areas for APT-FIM improvement and future research.

5.2. APT-FIM Computational Risk Requirements

The APT-FIM use cases highlighted three additional computational requirements and one communication requirement. The eight APT-FIM reactive metrics had intuitive applications but required scales on which to decide margins of error, tolerance, and budget constraint impacts. The core relationship measured by the model is the degrees of uncertainty any evidence has to the homomorphic objects of the APT reactive definition (Tab.1 elements 4-11). Trust in the model outcomes is enhanced by repeatability in the same conditions, and user confidence that correct judgements are made. The following sub sections clarify measuring uncertainty, assigning trust probabilities to non-homomorphic objects, and the impact of budget constraints on big data evidence processing.

5.2.1 Degrees of uncertainty when P(X)=1

The APT-FIM discrimination mechanism functions by comparing computer network secondary evidence to the eight

reactive homomorphic descriptions of Tab.1 and corresponding methods in Tab.2. A binary decision is made in step 1 to determine APT fit, yes or no, to each metric. However, uncertainty exists in any judgement and potential error margins require assessment. A judgement may have a true positive (TP), a true negative (TN), a false positive (FP), or a false negative (FN). The metrics from the confusion matrix set theoretical limits on certainty, and expectations for how many confirmations satisfy an acceptable risk assurance. In addition, in an investigation the critical resource is time, and the investigation budget defines the units of time for each process and related activities. Hence, the potential to locate confirmations is also controlled by budget. To set these limits the APT-FIM user must scope processes and activities that are both acceptable risk and acceptable budget resource allocation for the investigation. Hence, the question of how many confirmations is required to affirm a metric is governed by the trade-off of failure risk versus allocated budget units. With acceptance of a large failure risk and small budget unit allocation, one confirmation is enough but reducing the potential judgement failure rate requires increasing the confirmation number and subsequent budget unit allocations. In a big data context, a higher margin of error and fewer confirmations can enhance efficiency and not reduce the effectiveness of an investigation.

The second requirement from the use cases is to have guidance on how many of the eight metrics must be affirmative to distinguish an APT attack in reactive evidence. Each attack description for a metric may relate to a non-APT attack, and hence the eight metrics form a coherent set in which all must be affirmed for the highest assurance and approximation to P(X)=1. Then, all eight metrics require affirmation otherwise the evidence concerned triages to step 2 and is assigned a trust probability as defined in section 5.2.2. Cost efficiencies are also gained if any metric has no confirmation, then it is not necessary to continue expending resources on the other metrics and the investigation method moves to step 2. The logic behind the APT-FIM is to systematically reduce data processing by setting the trade-off curves to optimize expectations for acceptable outcomes. In each investigation it is the investigator's responsibility to set the appetite for risk and to maintain compliance with the budget controls.

5.2.2 Degrees of uncertainty when P(X)<1

When evidence fails acceptance at step 1 for P(X)=1 then it moves to step 2 where a trust probability is assigned. This sub section describes the controls and metrics for assigning trust probabilities in the model and the economic limits for information processing in big data contexts. In sub section 5.2.1 if sufficient evidence to satisfy the theoretical definition of APT is found then the investigation terminates for reporting. However, if insufficient evidence is found at step 1 then the investigation must continue to assign trust probabilities to data on a scale from 0 to <1 to represent the degrees of uncertainty for similarity. These trust values classify data for ranking, and the economic management of data volumes. Data with potential for a low trust probability is triaged to non-APT and the evidence with higher values retained. The investigator sets the cut-off trust probability value to fit the investigation appetite for risk and budget. Hence, evidence above the cut-off is included in the Report with the caveat of the trust probability measure. The probability tolerance mechanism further moderates efficient resource use for optimal effectiveness.

The structural design of the APT-FIM sets conditional weightings for trust probabilities by design. Computer network processes and activities layers may have APT evidence, but it is difficult to get and expensive to detect in large data volumes. The Layer 1 non sub class objects may be used by APTs but are not APT objects by binary design. Hence, objects in these two classes have low trust probabilities and low priority in an investigation. Objects in the metrics class are ambiguous for APT definition as they rely on the correct identification of an APT prior to reporting behaviors, time sequences, and so on. These objects are useful for investigation but have potential errors in the data methods, confusion of APT influences with non-APT influence, and metric tool limitations. Hence, by design the APT-FIM structure assigns low trust probabilities to network process and activities data, variable low to mid-range probabilities for metrics and semantics [Tab.5]. In these ways data volumes are reduced and investigation resources targeted to the highest probabilities for success.

TABLE 5
APT-FIM Structural Trust Probabilities

Structural Component	Trust Probability
Layer 1 non sub class	0%
Layer 1 sub class	100%
Layer 2 Semantic	0-50%
Layer 3 Network Process	0-10%
Layer 4 Network Activity	0-10%
Layer 5 Network Metric	0-50%

The APT-FIM semantic discrimination mechanism in Layer 2 applies eight metrics to determine APT evidence. The outcome of 8/8 is accepted in Step 1 however, the remaining possibilities require a weighted scale of trust probabilities. For example, 7/8 indicates disproportionally greater coherence with the APT definition than 6/8 and similarly the lower numbers with less coherence and value. The outcome of 0/8 may be assigned a probability >0 depending in which APT-FIM structural layer the evidence was collected. The application of the probability cut-off mechanism can eliminate large volumes of data based on where it comes from in the computer network and the probability value. Data reduction is critical to efficient and effective investigation in big data contexts. The APT-FIM prioritizes, ranks, and reduces the amount of evidence to be processed, when the investigator sets risk and tolerance values.

5.2.3 Degrees of uncertainty from budget constraints

The extent resources may be used in any investigation is determined by the budget. The investigation budget is a control that assures an outcome delivery and cost efficiency. In a reactive APT investigation, each process has budget constraints and in large computer network investigations it assures control of processes and related activities for cost efficiency. The impact of budget constraints is for the selection of efficient and effective methods, but also for the discovery outcome. An investigation is terminated when the budget constraints are satisfied introducing uncertainty as to the full discovery of evidence or related relationships. Hence, the budget can deliver

an indeterminate conclusion for reporting. The budget constraints drive urgency in the discovery processes to prioritize high value information in relation to the investigation objectives. The APT-FIM forces an economical outcome to any APT investigation by efficiencies when looking for degrees of uncertainty between the APT reactive definitions and any evidence in the investigation. Reductionist logic prioritizes the most likely evidence for decision making, reduces the amount of data for processing, and manages costs.

The investigation budget sets constraints for the expected expenditure of resources but relies on the investigator to set tolerance margins levels for optimizing compliance. A tolerance interval provides limits in which the budget is satisfied but there is discretion regarding the expenditure of resources. In practice, savings gained on one task can be reinvested into another task which has become more expensive, improving the quality of the investigation. The variation managed in the tolerance interval mitigates risks adopted when setting the appetite for risk and expected sample variations. The investigator can only work with samples of data, and Reports containing conclusions drawn from big data sample data, hence the setting of a confidence interval allows budget compliance and measures for uncertainty to report. For example, in big data any data is a sample from the complete population hence a confidence level can be set, such as a K score from statistical tables, to provide a measure for the accepted risk and caveat to Report conclusions. Budgets estimate and determine the extent resources may be used to meet investigation objectives and force a conclusion when satisfied.

5.3. APT-FIM Practitioner Use Requirements

The aim of an investigation is to deliver a trustworthy result while optimizing budget compliance with small data sets and acceptable risk levels. The APT-FIM achieves the performance expectations in two steps. Step 1 requires affirmative answers to eight questions, in Step 2 trust probabilities are assigned to evidence based on the structural network evidence location and the action of the APT-FIM semantic discrimination mechanism. All processes and related activities in both steps are in budget control for cost efficiency and effectiveness. The investigator must set a risk appetite to mitigate theoretical errors in decisions at Step 1, and report potential points of failure. In Step 2 tolerance bounds are set to control the amount of evidence to be processed before budget satisfaction. The calculation of budget and the related resource expenditure determines how much evidence can be processed before a conclusion is reported. The outcome of an investigation may be inconclusive but budget compliant.

Fig.6 summarizes the process steps in a flow chart an investigator is to take using the APT-FIM model. The use case learning and recommendations for improvement are included. The key additions for use are input data validation, setting budget units, budget compliance, and information disclosure. A robust determination of APT attack is then possible to a higher standard than many of the publications reviewed which only selected some APT features for decision making. The determination of an APT attack indicates that other attacks on network are probable from within the Green Zone, and specialized remediation actions are required to protect the computer network.

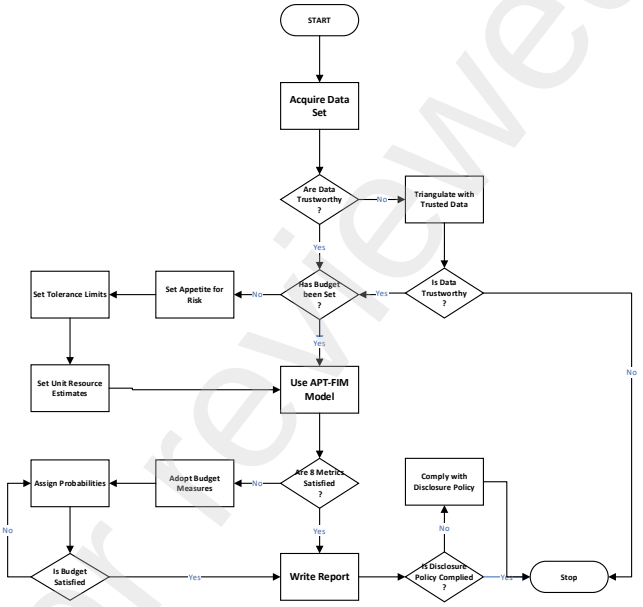

Fig.6 APT-FIM Use Chart

VI. Conclusion

The APT-FIM does not replace other tools and techniques that are used in network investigation but rather it adds a new tool that works with others in a computer network investigation. It differentiates APT evidence from competing representations by homomorphic comparison with the objects from the APT definition. Uncertainty is managed by first adopting theoretical measures for failure, and then systematically assigning probabilities to evidence where P(X)<1 in a controlled and cost-efficient discrimination +method. The APT-FIM increases capacity to effectively manage big data sets in large commercial networks by enforcing structures based on risk appetite, tolerance intervals, and budget constraints. The challenge for an investigator is to locate and identify evidence in relation to the investigation objectives from the network. To effectively perform investigation tasks then a range of tools are required. The APT-FIM is one that adds value in reactive discovery by strong evidential providence features and systematic data methods. Further APT-FIM research has been signaled in Section 5.1 arising from the use case analysis that concerns supervision, machine learning, and automation.

References

[1] X. Wenjun and R. Lagerstrom, “Threat modeling – A systematic literature review, J. Comp. & Sec., vol. 84, pp. 53-69, 2019.

[2] T. Zhu, et al., “General, Efficient, and Real-Time Data Compaction Strategy for APT Forensic Analysis”, IEEE Trans. Info. Forensics and Sec., vol. 16, pp. 3312 – 3325, 2021.

[3] L. Liras, et al., "Feature analysis for data-driven APT-related malware discrimination", J. Comp. & Sec., vol. 92, pp. 101660, 2021.

[4] G. Rattray, G., "Strategic Warfare in Cyberspace" (2001). MIT Press: MA., 1994.

[5] S. Hosseini, A. Jahangir, and M. Kazemi, "Digesting Network Traffic for Forensic Investigation Using Digital Signal Processing Techniques", IEEE Trans. Info. Forensics and Sec., vol. 12, pp. 3312 – 3321, 2019.

[6] C. Wang, et al., "Shrinking the Semantic Gap: Spatial Pooling of Local Moment Invariants for Copy-Move Forgery Detection", IEEE Trans. Info. Forensics and Sec., vol. 8, pp. 1064 – 1079, 2023.

[7] I. Friedbeig, F. Skopik, G. Settanni, and R. Fiedler, "Combating advanced persistent threats: From network event correlation to incident detection", J. Comp. & Sec., vol. 48, pp. 35-57, 2015.

[8] M. Auty, "Anatomy of an advanced persistent threat", J. Network Sec., vol. 4, pp. 13-16, 2015.

[9] N. Mohamed, and B. Bahari, "SBI Model for the Detection of Advanced Persistent Threat Based on Strange Behavior Using Credential Dumping Techniques", IEEE ACCESS, vol. 9, pp. 42919-42932, 2021.

[10] NIST (2016). NIST- SP 800-39 Managing Information Security Risk. DOI: https://csrc.nist.gov/publications/detail/sp/800-39/final

[11] N. Hoque, et al., "Network Attacks: Taxonomy, tools, system", J. Network and Comp. Apps., vol. 40, pp. 307-324, 2014.

[12] D. Ferraris, C. Fernandez-Gago, J. Daniel, and J. Lopez, "A Segregated Architecture for a Trust-based Network of Internet of Things", IEEE Con. Coms. & Networking Conference 2019, 2019. http://doi.org/10.1109/CCNC.2019.8651703

[13] A. Abass, et al., (2017). "Evolutionary Game Theoretic Analysis of Advanced Persistent Threats Against Cloud Storage", IEEE ACCESS, vol. 5, pp. 8482-8491, 2017.

[14] L. Zhang, et al., "A Game-Theoretic Method for Defending Against Advanced Persistent Threats in Cyber Systems", IEEE Trans. Info. Forensics and Sec., vol. 18, pp. 1349 – 1364, 2023.

[15] A. Abass, et al., "Evolutionary Game Theoretic Analysis of Advanced Persistent Threats Against Cloud Storage", IEEE ACCESS, vol. 5, pp. 8482-8491, 2017.

[16] Q. Zhui, and S. Rass, "On Multi-Phase and Multi-Stage Game-Theoretic Modeling of Advanced Persistent Threats", IEEE ACCESS. vol. 6, pp. 13958-13971, 2018.

[17] L. Huang, and Q. Zhu, "A dynamic games approach to proactive defense strategies against Advanced Persistent Threats in cyber-physical systems", J. Comp. & Sec., vol. 89, pp. 101660, 2020.

[18] A. Sharma, et al., "Orchestration of APT malware evasive manoeuvers employed for eluding anti-virus and sandbox defense", J. Comp. & Sec., vol. 115, pp. 102627, 2022.

[19] J. Jafarian, and A. Niakanlahiji, A. "MultiRHM: Defeating multi-staged enterprise intrusion attacks through multi-dimensional and multi-parameter host identity anonymization", J. Comp. & Sec., vol. 124, pp. 102958, 2023.

[20] W. Tian, et al., "Prospect Theoretic Study of Honeypot Defense Against Advanced Persistent Threats in Power Grid", IEEE ACCESS. vol. 8, pp. 64075-64085, 2020.

[21] M. Li, et al., "Study of APT Attack Stage Model", IEEE ICIS Conference, June 26-29, Okayama, Japan, 2016.

[22] M. Amin, et al., "Hidden Markov Model and Cyber Deception for the Prevention of Adversarial Lateral Movement", IEEE ACCESS, vol. 9, pp. 49662-49682, 2021.

[23] K. Horak, "Optimizing honeypot strategies against dynamic lateral movement using partially observable stochastic games", J. Comp. & Sec., vol. 87, pp. 101579, 2019.

[24] P. Zambrano, et al., "Technical Mapping of the Grooming Anatomy Using Machine Learning Paradigms: An Information Security Approach", IEEE ACCESS, vol 7, pp. 142129-142148, 2019.

[25] M. Au, et al., "Privacy-preserving personal data operation on mobile cloud—Chances and challenges over advanced persistent threat", J. F. Gen. Comp. Sys., vol. 79, pp. 337-349, 2017.

[26] S. Dara, S. Zargar, and V. Muralidhar, V. "Towards privacy preserving threat intelligence", J. Info. Sec. and Apps., vol. 38, pp. 28-39, 2018.

[27] J. Joloudari, et al., "Early Detection of the Advanced Persistent Threat Attack Using Performance Analysis of Deep Learning". IEEE ACCESS, vol. 8, pp. 186125-186137, 2020.

[28] A. Lemay, J. Calvet, F. Menet, and J. Fernandez, "Survey of publicly available reports on advanced persistent threat actors", J. Comp. & Sec., vol. 72, pp. 26-59, 2018.

[29] M. Kida, and O. Olukoya, "Nation-State Threat Actor Attribution Using Fuzzy Hashing", IEEE ACCESS, vol. 11, pp. 1148-1165, 2023.

[30] A. Villalón-Huerta, H., Marco-Gisbert, and I. Ripoll-Ripol, "A Taxonomy for Threat Actors' Persistence Techniques", J. Comp. & Sec., vol. 121, pp. 102855, 2022.

[31] J. Chen, C. Su, K. Yeh, and M. Yung, "Special Issue on Advanced Persistent Threat", J. F. Gen. Comp. Sys., vol. 79, no. 1, 243-246, 2018.

[32] C. Liu, and W. Chen, "The Study of Using Big Data Analysis to Detecting APT Attack", J. of Comp. 30(1), 206-222, 2019.

[33] H. Mookherjee, A., Theriault, and R. Wright, "A baseline for unsupervised advanced persistent threat detection in system-level provenance", J. F. Gen. Comp. Sys., vol. 108, pp. 401-413, 2020.

[34] J. Liu, et al. "Two statistical traffic features for certain APT group identification", J. Info. Sec. and Apps, vol. 67, pp. 103207, 2022.

[35] W. Niu, et al., "Uncovering APT malware traffic using deep learning combined with time sequence and association analysis", J. Comp. & Sec., vol. 120, pp. 102809, 2022.

[36] M. Khosravi, and B. Ladani, "Alerts Correlation and Causal Analysis for APT Based Cyber Attack Detection", IEEE ACCESS, vol. 8, pp. 162642-162256, 2020.

[37] I. Ghafir, *et al*., "Hidden Markov Models and Alert Correlations for the Prediction of Advanced Persistent Threats", IEEE ACCESS, vol. 7, pp. 98508-98520, 2019.
[38] B. Stojanovi, K., Hofer-Schmitz, and U. Kleb, "APT datasets and attack modeling for automated detection methods: A review", J. Comp. & Sec., vol. 92, pp. 101734, 2020.
[39] S. Singh, et al., "A comprehensive study on APT attacks and countermeasures for future networks and communications: challenges and solutions", J. Supercomp., vol. 75, pp. 4543–4574, 2019.
[40] C. Tankard, "Advanced Persistent threats and how to monitor and deter them", J. Network Sec., vol. 8, pp.16-19, 2011.
[41] A. Lajevardi, and M. Amini, "A semantic-based correlation approach for detecting hybrid and low-level APTs", J. F. Gen. Comp. Sys., vol. 96, pp. 64-88, 2019.
[42] P. Li, and X. Yang, X. "On Dynamic Recovery of Cloud Storage System Under Advanced Persistent Threats", IEEE ACCESS. vol. 7, pp. 103556-103569, 2019.
[43] Z. Mai, Q. Li, and X. Meng, "Discovering Suspicious APT Families Through a Large-Scale Domain Graph in Information-Centric IoT", IEEE ACCESS, vol. 7, pp. 13917-13926, 2019.
[44] L. Yangi, *et al.* "Security Evaluation of the Cyber Networks Under Advanced Persistent Threats", IEEE ACCESS, vol. 5, pp. 20111-20123, 2017.
[45] G. Zhaoi, K. Xu, L. Xu, and B. Wui, "Detecting APT Malware Infections Based on Malicious DNS and Traffic Analysis", IEEE ACCESS, vol. 3, pp. 1132-1143, 2015.
[46] A. Zimba, H. Chen, and Z. Wang, "Bayesian network based weighted APT attack paths modeling in cloud computing", J. F. Gen. Comp. Sys., vol. 96, pp. 525-537, 2019.
[47] D. Kao, "Performing an APT Investigation", IEEE 39th Annual International Computers, Software & Applications Conference, DOI10.1109/COMPSAC.2015.10.
[48] S. Wang; et al., "THREATRACE: Detecting and Tracing Host-Based Threats in Node Level Through Provenance Graph Learning", IEEE Trans. Info. Forensics and Sec., vol. 17., pp. 3972 – 3987, 2022.
[49] Z. Zulkefli, and M. Singh, "Sentient-based Access Control model: A mitigation technique for Advanced Persistent Threats in Smartphones", J. Comp. & Sec., vol. 51, pp. 102431, 2020.
[50] H. Sun, et al., "Impulsive Artificial Defense Against Advanced Persistent Threat", IEEE Trans. Info. Forensics and Sec., vol. 18, pp. 3506 – 3516, 2023.
[51] Misiko, N. "A Review of Metrics for Object Oriented Design", Global Journal of Computer Science, 20(2), 1-5, 2020.
[52] Whitmire, S. "Object Oriented Design Measurement", Wiley, New York, 1997.