International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 8 5259 - 5262

A Hybrid Methodology Approach for Fraud Detection Using Event Correlation Approach

Onyemauche U.C.
Department of Computer Science
Nnamdi Azikiwe University Awka
Anambra State, Nigeria.
Osigwe.uchenna@yahoo.com

Mbanusi, C.E .
Department Of Computer Science
Nnamdi Azikiwe University Awka
Anambra State, Nigeria.
Chary4sam@yahoo.com

Nwosu Q.N.
Department Of Physical & Health Education, University Of Nigeria
Nsukka, Enugu State, Nigeria.
ccnwosu@yahoo.com,
queen2010@yahoo.co

Abstract:- To effectively investigate mass of events oriented data, automated methods for extracting event records and then classifying events and patterns of events into higher level terminology and vocabulary are necessary. Semantically rich representation model and automated methods of correlating event information expressed in such models are becoming a necessity. The Event Correlation for Forensics (ECF) framework was developed with the strategic objective "to develop a means by which a consolidated repository of event information can be constituted and then queried in order to provide an investigator with post hoc event correlation. Key words: Semantics, Correlation, Digital Forensics.

*****

1. Introduction

Computer forensics emerged in response to the escalation of crimes committed by the use of computer systems and or other related digital devices either as an object of crime, an instrument used to commit a crime or a repository of evidence related to a crime. Digital forensics has been defined as the use of scientifically derived and proven methods towards the preservation, collection, validation, identification, analysis, interpretation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal or helping to anticipate the unauthorized actions shown to be disruptive to planned operations. One important element of digital forensics is the credibility of the digital evidence. Digital evidence includes computer evidence, digital audio, digital video, cell phones, digital fax machines to mention but a few.

Event Correlation refer to an array of technique applied to comprehending the dynamic behavior of system, based on events and patterns of events in their history. Garfinkel (2010) uses correlation techniques to identify similar features across entire corpuses of drive, a technique which could prove useful for identifying computers with similar usage pattern. Finally, another useful form of classification is similarity.

2. Related Research

Event correlation is term which has emerged from a number of computer security application domains, in particular in the Areas of network management and intrusion detection. It is used to describe an array of techniques applied to comprehending the dynamic behaviour of systems, based on events and patterns of events in their history. As these domains exist in the digital forensics domain, we find the need for event correlation.

Abbott et al (2011), have, in their Event Correlation for Forensics (ECF) research, translated textual log events into instances of a generalized data model (canonical form) implemented using a relational database to performing either interactive or automated scenario identification over these events. Stallard et al (2010), employed an anomaly based expert systems approach to identifying semantic inconsistencies in investigation related data. Their approach translated

MAC times generated by TCT and the UNIX last log into an XML representation, which was asserted into the HESS expert systems shell. Knowledge is encoded as heuristic rules which specify invariant conditions related to logins and potential file modifications.

Elsaesser et al (2006) employ an AI based approach to automated diagnosis of how an attacker might have compromised a system. Using a model of the topology of a network, the configuration of system, and a set of "action templates", a class of artificial reasoner called a "planner" generates hypothetical attack sequences which could have led to a particular situation. These hypothetical attack sequences are them run in a simulated environment, and the generated logs compared with the logs of the real world system. The action templates correspond to specifications of how a particular action will transit the state of the world from one state to the next.

Approaches to event correlation in the IDS and network management domains have focused on single domains of interest only, and have employed models of correlation that are very specific in nature. Repurposing these specific existing approaches to the more general task of event correlation in the CF domain is made difficult for a number of reasons. Existing event pattern languages do not necessarily generalize the application in wider domains. For example, while state machine based event pattern languages may work well for events related protocols, they do not work well with patterns where time and duration are uncertain. Most approaches focus exclusively on events, and ignore context related information such as environmental data and configuration information Furthermore, few approaches have available implementations in a form that is readily modifiable.

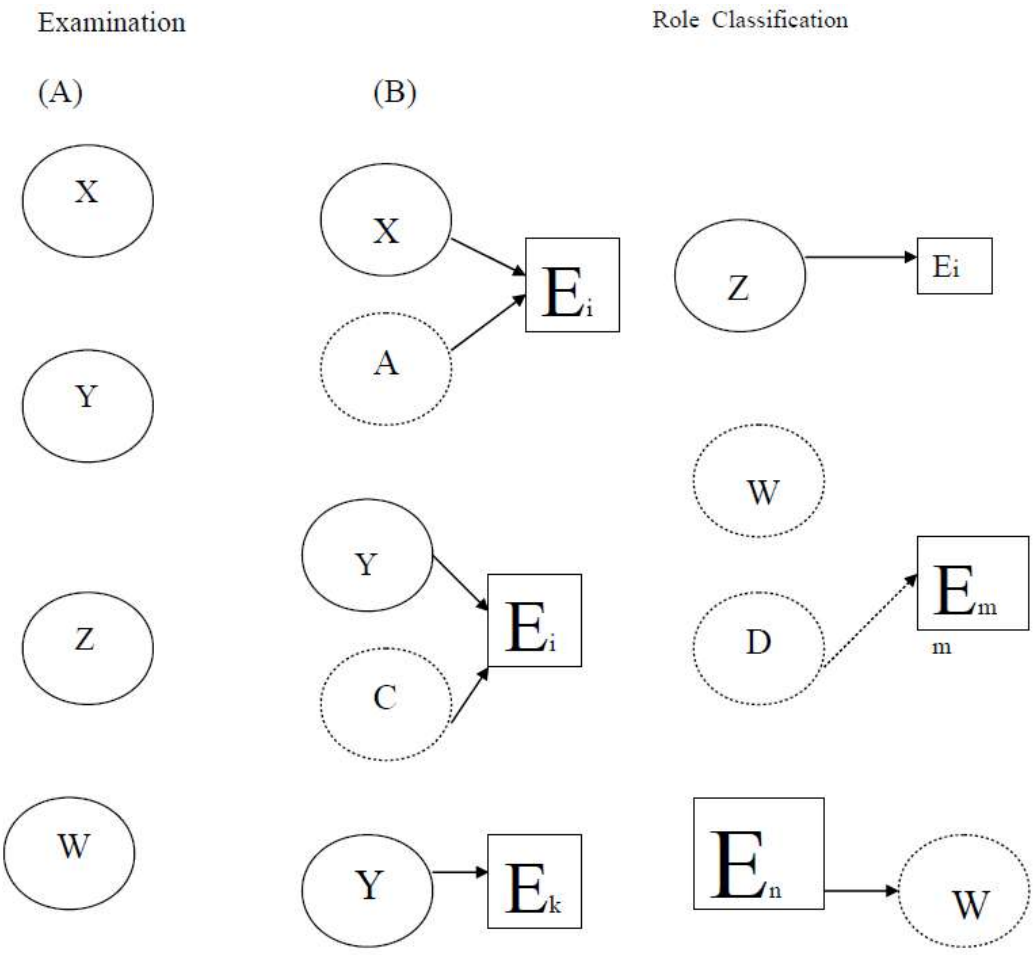
Where we have modifiable implementations of event correlation systems, we find that extension is complicated by the software paradigm underlying its implementation, and that the systems are weak on semantics. Adding new vocabulary to the event language is slowed because of compilation and linkage overheads. Addition of concepts outside of the event pattern language require reengineering of the STATL LANGUAGE compiler and supporting framework.

Kruse(2008) opined that the representation used to model events has a significant impact on the usability of correlation approaches, including conceptual expressiveness, extensibility, ease of integration of new information and maintainability. The MODEL language, a component of the DECS network management system, used an object oriented (OO) style model of classes of events related together in class/subclass relationship (which in this case was referred to as semantic generalization) . The event correlator translates from events patterns specified in the MODEL language directly to **C**++, and presumably, is encumbered by the maintainability characteristics of **C**++ software development and deployment. He further stated that Expert systems based approaches such as the EMERALD IDS combine a similar knowledge model, which support class/subclass models of events, with a rule language. The model however is dynamically constructed at run time, eliminating the **C**++ compile-link phase, resulting in simpler extensibility and more rapid evolution compared to the DECS approach.

A number of challenges were identified with the Event Correlation Forensics approach (ECF). The approach does not incorporate notions such as semantic generalization in its modeling approach does not identify a methodology for mapping the detail, rich domain specific information contained in log files to the canonical form implied that every event was seen as a time-subject-object-action tuple (TSOA), a notion which proved to be an impediment when attempting to represent arbitrary event log entries. This canonical form was supplemented by the addition of shadow data an arbitrary set of name-value pairs which could be associated with a canonical entry.

 ISSN: 2321-8169
5259 - 5262

3. Knowledge Representation



Evidence Classification and Heuristics Rule Classification

Where X, Y,W, Z are Objects, all E's are events and A,C,D are Targets After steps have been taken to preserve the state of the digital objects at the digital crime scene, the crime scene is searched for evidence. The goal of this phase is to recognize the digital objects such as X,Y,W and Z(digital computers) that may contain information about the incident. The first and foremost thing is to define a target that will be used to locate the evidence.For example, if you are looking for a file named foo.txt, then the target would have a name of foo.txt. If you are looking for a file with "bar" in the content, then the target would have "bar"in the content. Next is to extract data from the crime scene in some search pattern and then compare the extracted data with the target. After new evidence is found, updates of the general knowledge about the investigation so much recruited will be defined and reconstruction takes place giving birth to events (EK, Em, Ei).

4. Results

After all of the objects have been examined and their possible roles defined, event construction and testing groups the roles together to form events. Cause and effect roles are grouped together and if other objects must exist for the event to occur then they are searched for. The search may involve the objects that have been collected or it may involve a new search of the crime scene, if it is still available. After possible events have been constructed there may be objects that should exist, but could not be found. Hypotheses about the location of these objects are formulated.

5. Algorithm

1: SET Initiator identify initiator of the piconet
2: SET Target identify the target
3: SET MIN minimum number of agents to form a piconet
4: SET Agents[MIN] create an array of size MIN, index starts at 1g
5: for i = 1 to i _ MIN do
6: Entity[i] searchRange(Initiator) select an entity at random
7: NeighbourList getNeighbourList(Entity[i])
8: for j = 1 to j _ sizeof(NieghborList) do
9: for x = 1 to x _ MIN do
10: Entity[i].Suitable sendProbe-SW(NeighbourList[j], Target, Agent[x])
return TRUE if the Target and all agents are neighbours to Entity[i]
11: end for
12: if Entity[i].Suitable = TRUE then
13: sendT-SW(Entity[i])
14: Agent[i] Entity[i] fRecruit Entity[i]g
15: break Recruit another agent
16: else
17: continue flop and try different neighbors
18: end if
19: end for
20: end for

6. Conclusion

Event correlation is an activity which can be used to characterize activity on a computer system or systems. As such, it has significant value to Network forensics, which is a methodology conceived for the automated reconstruction of a computer system in order to provide direction for digital forensic investigations. By querying using heuristics rules it is possible to correlate the disparate sources of events and reconstruct a timeline of application or user activity in any computer system that gets registered to the host computer.

7. References

[1] Abbott, A.L, Beebe, R.E.(2011). A cyber Forensics Ontology: Creating a new approach to studying cyber forensics. Proceedings in the 6th Digital Forensics Research Workshop. Lafayette, IN. pp 11-13.

[2] Elsaesser, K.P, Gbendo, A.P.(2006). Dealing with Terabyte Datasets in Digital Investigations, Journal of *Research Advances in Digital Forensics*, Norwell: Springer, pp. 3-16.

[3] Garfinkel, H, E.(2010). Unifying Computer Forensics modeling approaches. A software engineering perspective. Proceedings in the 1st international workshop on systematic approaches to digital forensics engineering. Pp 10-15.

[4] Kruse, F.P.(2008). Overcoming Reasonable Doubt in Computer Forensic Analysis. A handbook on Digital Forensics evidence representation. Vo1 2Pp 12-16.

[5] Stallard, L.P, Peter, M.K.(2010). Computer Profiling to Assist Computer Forensic Investigations, presented at RNSA Security Technology Conference, Canberra, pp. 34-30.