e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:06/Issue:10/October-2024 Impact Factor- 8.187 www.irjmets.com

COMPARATIVE MULTIDATA FUSION NETWORK FORENSIC ANALYSIS PHASE FRAMEWORK FOR MANAGING SECURITY INCIDENTS

Mr. Peter Kiprono Kemei*1, Dr. Joel Cherus*2, Dr. Moses Thiga*3

*1,3Kabarak University Information Technology School Of Engineering And Technology, Kenya.

*2Forensic Security Consultant Expert Researcher, Kenya.

DOI : https://www.doi.org/10.56726/IRJMETS62234

ABSTRACT

Network forensics determines and retrieval of evidential evidence in a computer networked environs about a criminal activities which is admissible by grieved party. Computer forensic and data science field lays a robust foundation for network forensics as security frameworks, tools and techniques are in place for detecting, collecting, preserving and presenting breached information. Nevertheless, less has been done in mitigating phase analysis challenges from existing network forensic framework. The multidata fusion, data redundancy and integration evidences from various network sensors tools is the main challenge in analysis phase. The objectives of the study were to; analyse, investigate, identify, develop and evaluate a network forensic framework which addresses the multidata fusion, data redundancy and integration. A methodology was specifically formalized on real time and post attacked network traffic investigation based on datasets prototype implementation. The proposed technique in analysis phase is multidata fusion, data redundancy and integration traced datasets. The multidata fusion frameworks consolidates captured evidences from various network security sensors. The data redundancy algorithm eliminates data duplication and integration algorithm consolidate various attacked evidences into single entity attacks dataset.

Keywords: Network, Forensic, Framework, Analysis, Multidata.

I. INTRODUCTION

The analysis framework built by aggregating the strengths of open source tools in accomplishing the task of collection and analysis. Network security sensors with self-contradictory and corresponding utilities are used Security tools with similarity build the redundancy and reliability of attack information. Diversity among the tools will ensure versatility. Data fusion performed on the alert and attack information generated by these sensors so that the decision to ascertain the accuracy of the intruder data. In case suspicious packets are detected security sensors gives notifications inform of alerts. Packet capture and analysis tools or sniffers identify sessions or connections with anomalies in network traffic. Traffic statistics are read from packet captures or from Netflow records taken from the network connection through a monitored device. Network security sensors analyses the packet attacks in order to increase the confidence level of evidence. The redundancy and data integration algorithm validates sample data set with attack packets generated in desk check test in order to eliminate data duplication and consolidate into single entity. The accuracy of these tools is validated using confusion matrix theory or criteria of accuracy and False Alarm Rate (FAR) metric applied to measure the performance validation before making crucial decision to proceed with investigation.

Background to the Study

Analysis phase determine significance, reconstruct fragments of data and draw conclusions based on evidence found. There are numerous forms of exploitation that analysis phases (Aymen, 2020). (Dalal, 2021), network forensic information can also convey insufficiencies in the current sensors network security tools. These challenges makes existing analysis phase network forensic framework not able to present and provide admissible evidence. Network security tools can be applied to network traffic and data fusion performed on various output values generated. Many models have been proposed for data fusion of intrusion detection data. They are mostly based on confusion matrix theory of evidence. The models are surveyed to obtain the direction for data fusion in the context of network forensic analysis. In the same area, (Tiffanie,2021) Machine learning or artificial intelligence for sensor data fusion model. Pieces of evidence from heterogeneous defence systems are fused or combined to detect the attacks for anomaly detection and localization. Yuan (2021). Data fusion techniques effectively combine evidence from multiple sensors or a single sensor used in multiple places. A multi-source fusion system that uses the DS technique to combine beliefs from multiple security tools is investigated. Steffen (2019) proposed Efficient Attack Correlation and Identification of Attack Scenarios based on Network-model that different types of NIDS. The set of alerts can be partitioned into different alert tracks. IDSDFM consists of alert correlation module, security estimation module, and management & control module. Two types of alert aggregation is done - alerts that make up an attack and alerts that represent the behaviour of a single attacker. Steffen (2019) distributed intrusion detection system based on data fusion method consists of two layers: lower layer and upper layer. The lower one consists of host and network based sensors, which collect local features, and differentiates easy-to-detect attacks. The upper layer is a fusion control center, which makes global decision on these events by adopting confusion matrix combination rule. Humayun (2020) reviewied cyber security threats and vulnerabilities systematic mapping study that highlighted a notion where situation security analysis is made to understand threats and vulnerabilities from intrusion alerts and take appropriate actions. The network security situation elements are analysed, data is fused, and correlation identified using collared petri-net and the confusion matrix theory of evidence network security situation awareness fuses data from tools of intrusion detection systems, virus detection system, firewall, net flow records to monitor the network for intrusions and predict course of action. The security events are pre-processed and situation assessment is done through correlation to gather information about the attacks. Hussein (2021) proposed a pattern recognition approach is applied to network intrusion detection based on the fusion of multiple classifiers. Each member of the classifier ensemble is trained on a distinct feature representation of patterns, then the individual results are combined using a number of fusion rules. Expert knowledge about the characteristics that distinguish attacks from normal traffic can be used to extract features based on content (payload), intrinsic (network connection information) and traffic features (statistics). This evidence is combined to produce a final decision. Bouyeddou (2020) recommended an advanced intrusion evidence automated analysis system framework that collects the evidences from integrated statistical approach based on cyber-attacks. Alerts of network intrusion detection systems are treated as primary evidences and logs from vulnerability scanner, network monitors, firewalls and others can be used as secondary evidences. The log collection agent collects intrusion logs from various sensors, pre-aggregates and adds a signature. Log collection agent collects has security event parsers that use regular expressions to automate log aggregation. An ERA is proposed for retaining effective information after removing redundant information. Analytical intrusion detection framework proposed for information integration and realization of a distributive intrusion detection systems environment with multiple sensors and a mechanism for selecting and integrating the probabilistic inference results to aid most probable forensic explanation. The probabilistic approach is also used for integrating information from different sensor sources in a distributive network intrusion detection systems environment. A novel cyber data fusion system (Sahu, 2021) proposed to specifically address the tracking and projection of multi-domain data fusion detection evidence attacks. It uses information fusion to provide situation awareness and threat prediction from massive volumes of sensed data. The system is based on information fusion engine for real-time decision-making and threat assessment of network data and information. Information fusion engine for real-time decision-making efficiently correlates IDS alerts, identifies individual multistage attacks, and provides situational measures of the identified attacks. Threat assessment of network data and information fuses information extracted from each attack track estimates, to determine threatened entities and differentiates them by threat scores. Sheikhalishahi (2022) proposed model based model that provides a technique of Privacy preserving data sharing and analysis for edge-based architectures. The proposed framework computes the best trade-off among privacy and result accuracy, based on the privacy requirements of data providers and the specific requested analysis algorithm.

II. METHODOLOGY

The attacks identified in the examination phase analysed by performing data fusion of information from multiple security sensors. Security sensors with complementary and contradictory functions are used. Security tools with similarity build the redundancy and reliability of attack information. Open source tools, commonly available on the network to gather attack information and generate alerts, attack statistics and information based on specific attacked information evidence. They are designed for active and post-mortem investigation of packet captures. Full packet data is captured by this tools, stored in a host and analysed off line at a later time. They also collect statistical information based on some criteria within the network traffic as it passes through the network. The network equipment collects this data and sends it to a flow collector which stores and analyses the data. These tools are able to trace packet based systems involving full packet captures at various points in the network. The packets are collected and stored for deep packet inspection by forensic investigators. The information produced by these tools in one stage are characterized and transformed for use by other tools in the succeeding stages. The data sets are partitioned, system are trained and then tested and integrated so that the investigator can have an edge over the attacker. These tools identify and tracks files across local area networks and across the World Wide Web, for the purpose of gathering intelligence and forensic evidence and standard means of extending the functionality of other tool architecture. Time consuming and error prone processes are identified and automated. These tools are used to converted attack packets back into the same format harnessing their strengths which enable analysis. Diversity among the tools ensures versatility description of network forensic analysis tools and description of Network Security and Monitoring (NSM). The main role of examination phase implement the diversity of network sensors techniques enhanced analysis of feature selection method where network traffic has many features to measure. The problem is that with the huge amount of network traffic we can measure many irrelevant features. These irrelevant features usually affect the performance of detection rate and consume the network security techniques resources, clustering and classification of network security incidents and data integration. These steps leads grouping of selected attacks to differentiate false positives, redundancies, true alerts and reducing the size of information to manageable level. The is achieved through multi-data fusion through implementation of combination of security sensors which enhance attacks evidence accuracy and manageable level for better unitization of network security resources.

III. MODELING AND ANALYSIS

The analysis framework is built by aggregating the strengths of open source tools in accomplishing the task of collection and analysis. Network security sensors with self-contradictory and corresponding utilities are used. Security tools with similarity build the redundancy and reliability of attack information. Diversity among the tools will ensure versatility. Data fusion is performed on the alert and attack information generated by these sensors so that the decision is more accurate. In case suspicious packets are detected by network intrusion detection systems givens notifications inform of alerts. Packet capture and analysis tools or sniffers identify sessions or connections with anomalies in network traffic. Traffic statistics were read from packet captures or from Netflow records taken from the network connection through a monitored device. The following network sensors tools were used in analysis phase framework architecture for multi data as shown in figure 1.

Fusion Snort: It is network intrusion detection systems open source software with ability of examining, detecting, analysing and logging network packet traffic during transit with suspicious features and contrary to define rule set configured by the user. Snort has also ability of decoding, printing logs, alerting and capturing full packets headers evidence messages by default. One of the features of snort is fast alert mode, which has ability to write, read and analysis packet file format detailing the alert massage, timestamp, source IP, destination IP and port numbers.

Wireshark: Wireshark is open source network software with ability of capturing and analysing real time network traffic during transmission. It has a module used to output packet information longest side protocol evidence. The sensors ability is to captures the packet where the forensic investigator can be a position to read, import, and export and saved the contents. It can also filter, search contents based on specified criteria and create numerous statistics specified by the investigator. Wireshark can be integrated with other network sensors in decoding protocols contents in order to dissector other high-level protocols as well.

Tcpstat is open source software able to monitor and report by reading certain network statistics interfaces. It has ability to read previous tcpdump stored file as well as calculate the packet traffic such as bandwidth, speed, packet average size, load of particular interface, standard deviation of packet size etc. It calculates statistics like bandwidth, number of packets, packets per second, average packet size, standard deviation of packet size, load of particular interface and ability to manage various transiting packets per second.

Bro is an open source UNIX based NIDS software capable of detecting and monitoring network traffic that has been manipulated and attacked by intruders based on contents and characteristic nature. It collects, filters, and analyses traffic that passes through a specific network location. Bro occur with fixed procedure code calculated for detection of most common intrusion of Internet application. These policies incorporate a signature matching facility that looks for specific traffic content. It can also analyse network protocols, connections, transactions, data amounts and many other network characteristics.

Basic Security Analysis Engine (BASE): It is originates from ACID code and has ability of offering front-end web interface for analysing and querying incoming alerts from other network security sensors such as snort. It has high sensing ability to detect attacks that snort cannot detect in case of intrusion. It is used to supplement other security sensors in a network forensic investigation through web interface module. It allows security investigator flexibility to make decisions based on what and how much each user can access information through authentication mode. We convert the contents of the pcap file into a database and reconvert the attack packet records in the database to a new pcap file using the Net::Pcap module of the Perl language. The file contents of lib-pcap captures packet information containing the protocol types used such as TCP, IP, UDP, ICMP and Ethernet alongside encapsulation details. These protocol features are extracted recursively from Lib-pcap file and inserted in to database table. Another lib extension file is Net: Pcap that can used to encode and extract protocol features. The main protocol attributes captured by Net. Cap file extension and extract include protocol ver, tos, hlen, id, offset, ttl, cksum, src_ip, dest_ip, tos, proto_type and options. These attributes are encapsulated, copied and stored into specific table in forensic database. These packets attribute creating header evidence of network protocols intruded by an intruder in compromising the user network end systems. A new packet capture file is created from the attack packets which is minimum in size and with maximum possible information as evidence. The pcap file with only attack packets is very much reduced in size when compared to the integrated file.

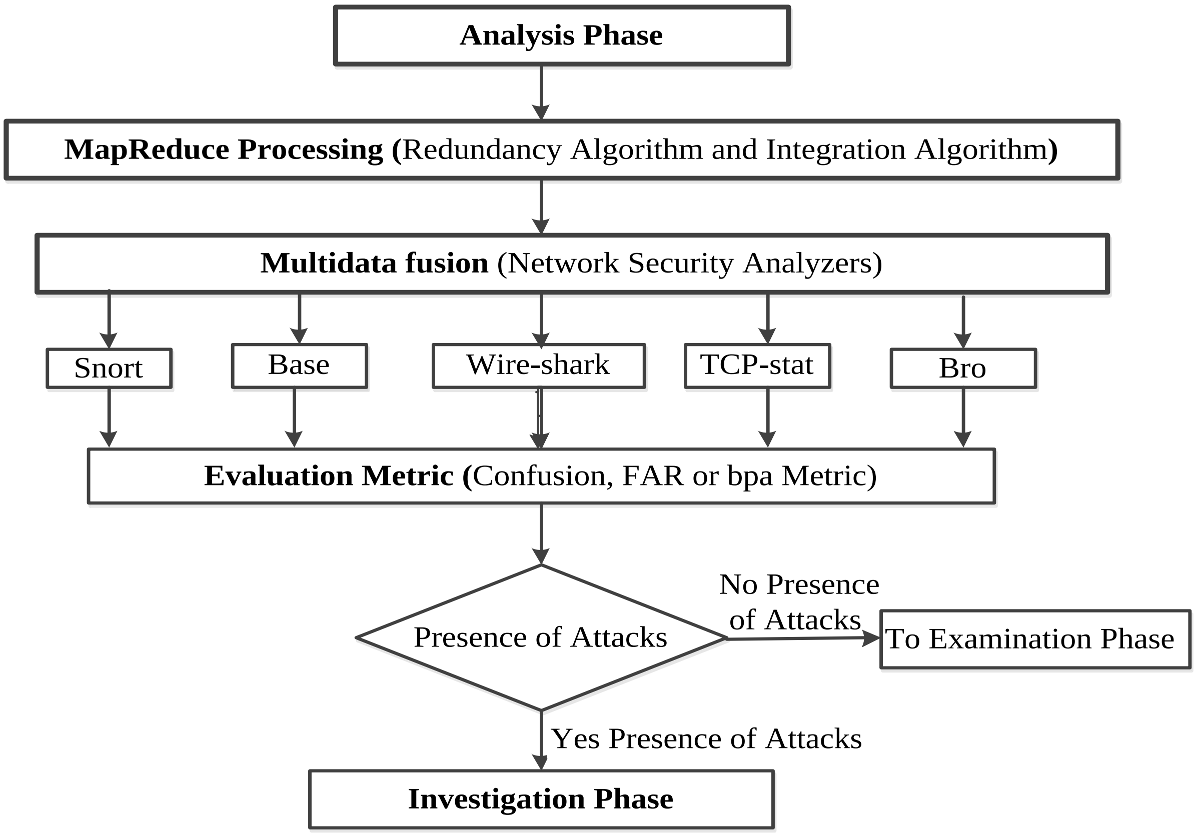

Analysis Phase
MapReduce Processing (Redundancy Algorithm and Integration Algorithm)
Multidata fusion (Network Security Analyzers)
Snort
Base
Wire-shark
TCP-stat
Bro
Evaluation Metric (Confusion, FAR or bpa Metric)
Presence of Attacks
No Presence of Attacks
To Examination Phase
Yes Presence of Attacks
Investigation Phase

Figure 1: comparative multidata fusion network forensic analysis phase framework for managing security incidents

IV. RESULTS AND DISCUSSION

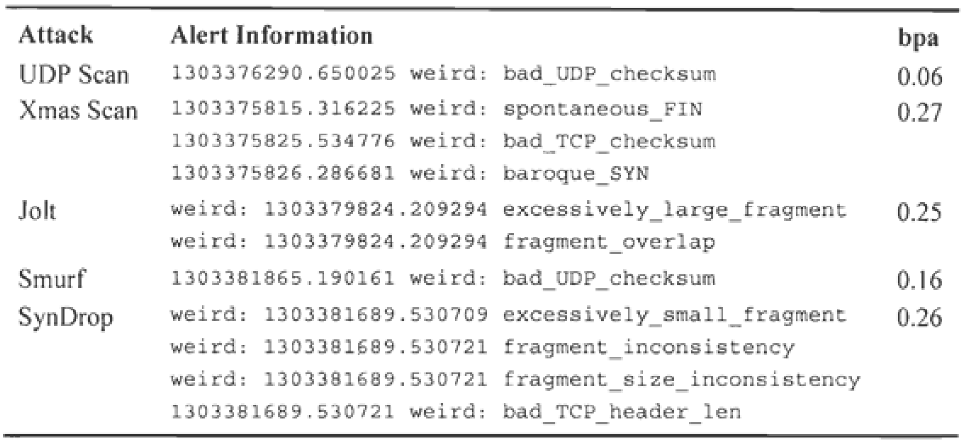
The intrusion dataset was captured by the network tools and codes. The confusion matrix theory or criteria of accuracy and False Alarm Rate (FAR) metric applied to measure the performance of the proposed scheme for analysing and tracing the attacked vectors of evidence for identification of the suspicious addresses detected by the multi fusion sensors tools applied to test the admissibility and validate the attacks evidence.Based on several combination of network sensors used in multidata fusion to analyse the evidences information captured to calculate the belief functions measure using the confusion matrix according to (Heydarian, 2020) multi-label confusion matrix rule of combination as given in equation below

$$\text{FAR}(m) = \frac{\text{FP+FN}}{\text{TP+TN+FP+FN}} \qquad (\text{ i})$$

TP represents true positive which denotes a number of the correctly attack classified, TN represents true negative expressing a number of the correctly normal classified, FP represents false positive the number of the misclassified attacks and FN represents false negative the number of the misclassified normal records.

FAR for the five attacks and using the five tools. The values are shown in Table 1 as given below.

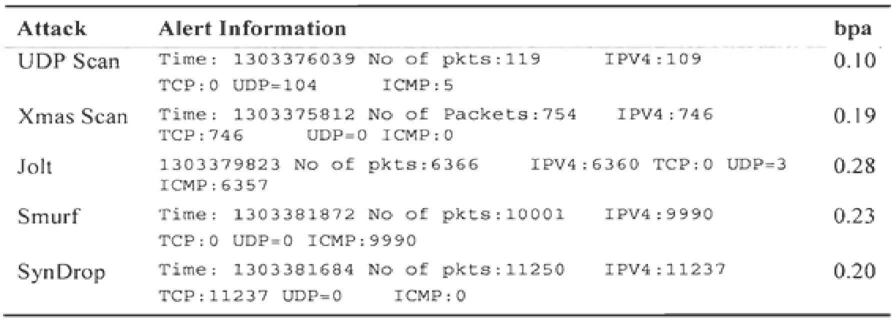
Table 1: Statistics Information Generated by Bro Sensor



Attack	Alert Information	bpa
UDP Scan	1303376290.650025 weird: bad_UDP_checksum	0.06
Xmas Scan	1303375815.316225 weird: spontaneous_FIN 1303375825.534776 weird: bad_TCP_checksum 1303375826.286681 weird: baroque_SYN	0.27
Jolt	weird: 1303379824.209294 excessively_large_fragment weird: 1303379824.209294 fragment_overlap	0.25
Smurf	1303381865.190161 weird: bad_UDP_checksum	0.16
SynDrop	weird: 1303381689.530709 excessively_small_fragment weird: 1303381689.530721 fragment_inconsistency weird: 1303381689.530721 fragment_size_inconsistency 1303381689.530721 weird: bad_TCP_header_len	0.26

Bpa's given for various statistic information generated by snort, Tcpstat, Base and Wireshark are shown in Table 2.

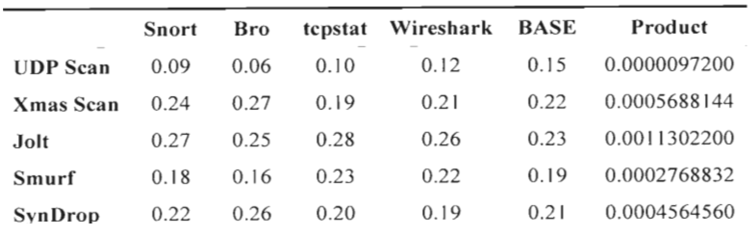
Table 2: Statistics Information Generated by Snort, Tcpstat, Base and Wireshark



Attack	Alert Information	bpa
UDP Scan	Time: 1303376039 No of pkts:119 IPV4:109 TCP:0 UDP=104 ICMP:5	0.10
Xmas Scan	Time: 1303375812 No of Packets:754 IPV4:746 TCP:746 UDP=0 ICMP:0	0.19
Jolt	1303379823 No of pkts:6366 IPV4:6360 TCP:0 UDP=3 ICMP:6357	0.28
Smurf	Time: 1303381872 No of pkts:10001 IPV4:9990 TCP:0 UDP=0 ICMP:9990	0.23
SynDrop	Time: 1303381684 No of pkts:11250 IPV4:11237 TCP:11237 UDP=0 ICMP:0	0.20

The probability of believing that network attack has been carried out and occurred is based on the degree of alert credible evidence used to launch attack. The evidence belongs to same set **FAR(m)** of attack information but they are considered to be special subset of **FAR(m)** information. The measure plausibility degree of alert information to ascertain the evidence belongs to set A of attack information or to any of its subsets or to any set that overlaps with **FAR(m)**. Based on several combination of network sensors used in multidata fusion to analyse the evidences information captured to calculate the belief functions measure using the confusion matrix rule of combination as given in equation (i).

Table 3: Combined Sensors Statistics information generated by snort, Bro, tcpstat, Wireshark and Base.



	Snort	Bro	tcpstat	Wireshark	BASE	Product
UDP Scan	0.09	0.06	0.10	0.12	0.15	0.0000097200
Xmas Scan	0.24	0.27	0.19	0.21	0.22	0.0005688144
Jolt	0.27	0.25	0.28	0.26	0.23	0.0011302200
Smurf	0.18	0.16	0.23	0.22	0.19	0.0002768832
SynDrop	0.22	0.26	0.20	0.19	0.21	0.0004564560

The calculations for a two attacks, UDP Scan and Jolt are shown in the table 16 above. The numerators are the product of all the bpa's for the Jolt and UDP Scan respectively. The denominators are the summation of product of all bpa's of the individual sensors.

The value of m for UDP Scan was calculated as shown below:

$$Bro\ UDP\ Scan\ (bpa\ or\ m) = \frac{0.06}{0.06 +\ 0.10}$$

$$= \frac{0.06}{0.16} = 0.375$$

The value of bpa or m for (combined security sensors) UDP Scan was calculated as shown below:

$$m(combined\ ) = \frac{0.0000097200 + 0.06}{00000972 + 0.09 + 0.06 + 0.1 + 0.12} =\ 0.11538$$

$$Ratio = \frac{Bro\ scan\ (bra\ or\ m)}{m(combined\ Scan\ UDP)} = \frac{0.375}{0.11538} = 3.2$$

The values FAR (m) in both sensors seem to be very low small when compared to the assigned values of FAR (m) since five sensors tools were considered for attack evidence data fusion. The value of m would be less than individual FAR (m) if only individual tools were put into consideration. Nevertheless, it is very insignificant from the calculation to put into consideration from the values of m that the UDP scan attack has transpired. The probability of m value of combined attack is three times less than m value of UDP scan attack based on Bro sensors. This implies that results attained from the calculation of m's values above illustrate that combination from combining many network sensors gives a strong confidence belief that the attack has taken place as compared to the belief of using only one individual network sensor. The same procedure or steps applied for other types of network attacks. From calculation above it illustrates that it is more accurate to prove and validate that an attack has taken place when attack evidence is fused and subjected to combination of network sensors tools. Based on evaluation criteria of accuracy or FAR (m) metric applied to measure the performance of the proposed network forensic framework for managing security incidents in examination phase while identifying and tracking security incidents. The beliefs of FAR metrics are dispersed by a forensic investigator manually stepwise and the values are determined based on the years of expertise of how to monitor the network traffic. This depends on evidences generated by the specific network sensor tool ability or specific attack occurrence security incidents launch by an attacker. The type of security sensors determines the value of “m” or threshold"τ” that is calculated using same procedure. When the value of “bra or m” is greater than the value of"τ” for a given combination of sensors used, then we can prove that indeed attack has occurred based on evidence. It the value of “m” is less than the value of"τ”for a given set of network sensors then we can prove that an attack has not occurred. This evidence validate that attack has taken place based on this information which can be relied upon and strengthen the forensic investigation before making informed decisions if an attack to place or not. We can use same network tools in ensuring there no data versatility and in cases of redundancy. This kindly of combining number of network sensors for data fusing analysing based evidence captured determines strong evidences of proving that an attack has been launched. This kind of open source network tools can established strong belief and evidences for analysing post mortem network attacks based on captured network traffic and files. The sensors security tools are setup and configured to capture the network logs traffic or plain text files for output analysis purposes. The logs or plain files are captured by sensors tools automatically whenever there is alert of attack traces information and validation of information proved by using combination confusion matrix metric theory criteria of accuracy and FAR (m) metric applied to measure the performance of the proposed scheme for analysing and tracing the attacked vectors of evidence. The additional evidences are negligible which is taken as overhead due to time spent during data fusion and time taken when combination of security sensors are capturing the attack information or giving alerts as compared to time taken when individual security sensors are used. Sensors relies alerts attacked network events evidence which are subjected to confusion matrix and FAR evaluation metrics to validate the evidence accuracy.

V. CONCLUSION

The analysis phase presents attacked packets and alerts, which are not admissible since no reconnaissance performed from various security sensors. The phase implements data fusion that alert and attack information generated by network security sensors so that the attack evidence presents accurate that ascertain the validity of the attack occurrence. It fuses the identified and validated multi data evidence from multiple security sensors. It also classifies attacked patterns using data mining, soft computing or statistical approaches. A technique for performing data fusion of information from multiple security sensors was identified, implemented; tools with complementary and contradictory functions were also identified. Data fusion was performed on the alert and attack information generated by these network sensors so that the attack evidence eliminates the redundancy and ascertain the accuracy of captured attacked evidence. Confusion matrix theory or criteria of accuracy and False Alarm Rate (FAR) metric applied were used to measure the performance of the proposed scheme for analysing and tracing the attacked vectors evidence perform fusion of the alert information and the validity of the attack occurrence was ascertained. The attack information detected and validated where crucial decision shall be to proceed with the investigation phase. The results achieved improve the expectation and reliability as compared to existing network forensic framework infrastructure.

VI. REFERENCES

[1] Aymen, A. M. F. S. Ontology-Based Smart Sound Digital Forensics Analysis for Web Services. Digital Forensics and Forensic Investigations: Breakthroughs in Research and Practice Volume 1, Issue 24. 8 April (2020).

[2] Dalal, M., Juneja, M. (2021). Steganography and Steganalysis (in digital forensics): a Cybersecurity guide. Multimed Tools Appl, Volume 3, Issue 8. 7 September, 2021 pp 5723–5771,

[3] Tiffanie, E. S. M. On Exploring the Sub-domain of Artificial Intelligence (AI) Model Forensics. Digital Forensics and Cyber Crime , Volume 2, Issue 11. 10 June (2021), pp 35–51.

[4] Yuan Zuo, X. Z. Heterogeneous big data fusion in distributed networking systems for anomaly detection and localisation. International Journal of Security and Networks, Volume 5, Issue 2, 2 March 2021, pp 220-229.

[5] Steffen, H. F. W. Efficient Attack Correlation and Identification of Attack Scenarios based on Network-Motifs. 2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC), , Volume 5, Issue 4, 20 Feb. 2019 pp1-11.

[6] Humayun, M. N. . Cyber Security Threats and Vulnerabilities: A Systematic Mapping Study. Computer Engineering and Computer Science, Volume 6, Issue 2, 30 July 2020 pp 3171–3189.

[7] Hussein, E. Proposed intelligence systems based on digital Forensics:. International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology. Volume 10 Issue 6 10th August 2021, pp. 32-42. Malaysia: Elsevier Ltd.

[8] Bouyeddou, B. H. Detecting network cyber-attacks using an integrated statistical approach. Cluster Comput 24. Cluster Computing, Volume 10, Issue 6 21 Nov 2020, pp1435–1453.

[9] Sahu, T. Multi-Source Multi-Domain Data Fusion for Cyber-attack Detection in Power Systems, in IEEE Access. Volume 10, Issue 6, 15 June 2021, pp 119118-119138.

[10] Sheikhalishahi, M. S. (2022). Privacy preserving data sharing and analysis for edge-bas ed architectures. Int. J. Inf. Secur. . International Journal of Information Security, Volume 1, Issue2 2022, pp 79–101.

[11] Heydarian, M. T. E. (2022). "MLCM: Multi-Label Confusion Matrix," in IEEE Access, Volume 4, Issue 10, pp. 19083-19095.