NETWORK FORENSICS: REVIEW, TAXONOMY AND OPEN CHALLENGES

Janitza Punto
(17512087)

Content

1. Introduction
2. The importance of network forensics
3. Current network forensics techniques
4. Comparison of network forensics techniques
5. Open challenges in network forensics
6. Conclusion and future directions

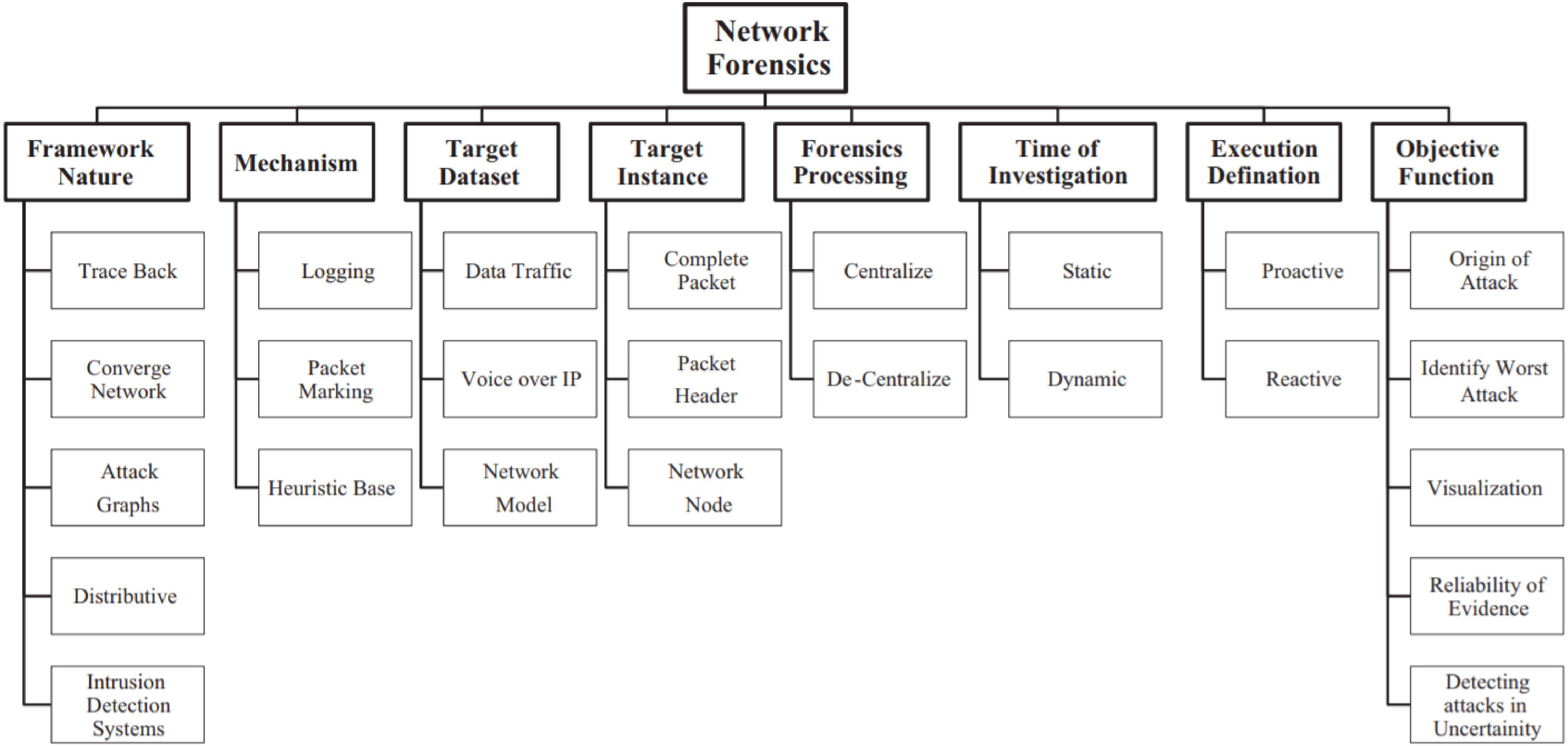
1.Introduction

- There were several investigation methods for network security breaches and vulnerabilities, which rely on identifying, capturing, discovering and analyzing network traffic encompassing network devices and infrastructure.
- Network Forensics permits to explore digital evidence in the network traffic after the occurrence of the suspected event.
- Traditionally, network forensics reconstructs network attack by capturing network traffic at one device and transmits it to other devices for analysis. However, this overloads the communication channel and generates time delays; and also, results in poor incident response.
- Refined methods are required for analyzing network traffic. Over the years an extensive range of network forensics techniques (NFT) has been proposed.
- This paper reviews the fundamental mechanics of NFTs, proposes a thematic taxonomy for the classification of current NFTs based on its implementation and target data sets, discusses similarities and differences in current NFTs, and finally deliberates about open research challenges in network forensics.

2. The importance of network forensics

- Organizations are concerned about their network and data security due to many attacks on different companies, such as DDoS attacks to social networks and phishing attacks. The criminals have to be traced out and legal evidence is required in the court to convict them.
- Companies can attract users towards their market portfolio by providing data and network security in e-transactions, e-business, and other Internet based services by increasing trustworthiness for users and ability to safeguard their interest. They have to monitor and analyze their network traffic to detect malicious events and deal with the attack as quickly as possible.
- In order to detect malicious packets or malicious programs, active monitoring of certain events is carried out. Techniques for active monitoring include anomaly detection, signature scan detection, intrusion detection systems, access control list and honeypots.
- Network forensics analyzes historical network data in order to investigate security attacks by reconstructing sequence of security attacks.
- Besides network attacks, network forensics is applicable to address network issues of business critical systems.

3. Current network forensics techniques



3. Current network forensics techniques

This section reviews frameworks of current NFT. Each NFT is illustrated in terms of:

- Objective
- Forensic Approach
- Methodology
- Detection of attack
- Characteristics
- Performance
- Critical aspect

3.1 Traceback based NFT

Traceback	Objective	Forensic approach	Methodology	Detection of attack	Characteristics	Performance	Critical aspect
TDPM	Traceback DoS attacks	Hash correction codes	Packet marking	Known attacks	Accuracy and scalability	Less false positive rates	Time consuming, lack of management module, application layer traffic is not captured.
CFS	Determine fake values in SIP request	Collaborative scheme	Network operator records	Unknown attacks	Time and storage efficient	Less time consuming in analysis	Non-identification of anonymous attacks, dependency on network operators and SIP registrar data
NFEA	Provides integrity to the collected evidence	Authenticated evidence and flow-based selection marking scheme	Packet marking	Known attacks	Minimize overhead of network throughput	Performance enhance when applied to selected packets	Not effective when attacker hides MAC address, not adopted for IPv6 address, edge router memory not enough, lack of scalability and reliability in large network systems.
LWIP	Trace DDoS attacks	Lightweight IP traceback	Packet marking	Known attacks	Path reconstruction	Significant path reconstruction	TTL can be not valid for analysis, only targets IPv4 headers.
Scalable-NF	Self-propagating attacks identification	Scalable based network forensics	Logging	Known attacks	Accuracy, space and time efficient	Capture real time traffic	Possibility of biased result, need of extra resources, lack of automation system to visualize real time data.
HB-SST	Identify attack in anonymous communication	Hopping based spread spectrum	Spread spectrum	Unknown attacks	Accuracy and secrecy	False positive rate decreases	No easily scalable, time consuming in large networks, used codes can be altered easily, can be affected when user changes frequency.
ITP	Real-time attack investigation	Real-time and periodic analysis	Logging, packet marking	Unknown attacks	Space efficient probabilistic data structure	Accurate attack detection, less false positive rate	System complexity, need of many computational resources in large networks, un-updated router attack pattern list.

3.2 Converge network based NFT

Converge Network	Objective	Forensic approach	Methodology	Detection of attack	Characteristics	Performance	Critical aspect
PBNF	Identification of attack patterns	Log correlation	Logging	Known attacks	Robust and flexible	Reduce false positive rate	Not easily scalable, storage not enough, forensic server may become bottle neck.
VoIP-NFDE	Identification of malicious packet in network traffic	Digital evidence with network forensics	Logging	Known attacks	Accuracy, storage efficient	Filtering of network traffic for analysis	Not easily scalable, lack of solution to capture dispersed voice packets, time consuming, huge storage resources required.
VoIPEM	Reconstructs attack path	VoIP evidence model	Logging	Known attacks/Unknown attacks	Integrity and reliability	Identification of attacks within less information	Does not identify attacks in anonymous communication, huge storage resources required, depends on strong hypothesis.

3.3 Attack graphs based NFT

Attack graph	Objective	Forensic approach	Methodology	Detection of attack	Characteristics	Performance	Critical aspect
SA	Identify attack and their impact on enterprises	Scalable analysis	Dependency graph	Known attacks/Unknown attacks	Measure current and future attacks	Efficient for small network	Lack of automatic generation of dependency graphs, lack of categorization in attack graph, lack of visualization interface.
AGFE	Monitor intruder actions	Anti-forensics	Forensics examination	Known attacks	High overall security, accuracy	Evaluate the alteration in traces	Less scalability, anti-forensics nodes should be in sensitive part of the network, big storage space required, system complexity, time consuming.
MLL-AT	Investigation of multi-level attacks	Multi-level and Layer attack tree	Network modeling	Known attacks	High accuracy	Determine system risk	Less scalability, big storage capacity required, time consuming, difficulty in weighting nodes in attack tree in case of DDoS attacks.
FCM	Identification of worst attack	Finite cognitive map	Genetic algorithm	Known attacks	Less complex	Less false positive rate	Interference of human knowledge, less scalability, lack of visualization interfaces.
CBSH	Root cause of the attack identification	Design model	Probabilistic approach	Known attacks	Adaptability, scalability	Complexity 0 (MN2)	Depends on human observation, high expertize required, time consuming, not adaptive in real-time network data traffic.

3.4 Distributive based NFT

Distributive	Objective	Forensic approach	Methodology	Detection of attack	Characteristics	Performance	Critical aspect
ForNet	Distributive analysis	Bloom filter tracking	Logging	Known attacks	High response time, Light weight filtering	Trustworthy information	Storage of raw network data, probability of undetected attacks, logs sent to forensic server in risk.
DCNFM	Integrity and validity of evidence	Mapping topology, network attack statistic	Logging	Known attacks	Classification, link and sequential analysis	Identify potential risk	Huge storage space required, logs might lose integrity, forensic server may become bottle neck, time consuming and complex in real-time situation.
DRNIFS	Real-time network intrusion analysis	Log and network traffic analysis	Logging	Known attacks	Robust, less false positive rate	Quick incident response	Huge storage capacity required, data integrity in risk, forensic server may become bottle neck, more consumed resources and delays due to encrypted communication.
DNF-IA	Integrity and authenticity for evidence	Multi-Immune theory	Logging	Known attacks	Scalable, high response	Real-time analysis of the attack	Non-scalable approach for forensic server in large network, forensic server may become bottle neck.

3.5 NFT using intrusion detection systems (IDS)

Intrusion detection system	Objective	Forensic approach	Methodology	Detection of attack	Characteristics	Performance	Critical aspect
AIDF	Identification of unidentified signature rule	Probabilistic model	Probabilistic inference	Unknown attacks	Scalability, extracting hide information	Prefect discovery, flexible, robust	Lack of knowledge base modules to store untreated hidden data
DFITM	Forensic server tolerance	Formal methods	Formal analysis	Known attacks	Separation of malicious traffic	Availability, high throughput, tolerance	Big storage capacity required, data security in risk due to storage in multiple places.
IIFDH	Monitoring log files	Steganography	Logging	Known attacks	Integrity and correctness for evidence	Integrity of evidence with real-time detection	Forensic server may become bottle neck, processing and storage problem in forensic server, large bandwidth required, extra security measures required for protecting important components of the network.
NFIDA	Credibility and reliability for evidence	Multi-dimensional analysis	Logging	Known attacks	Data encryption, multidimensional analysis	Capturing of complete network traffic	Lack of real-time network analysis, less scalability, newly collected network traffic in storage space overwrites existing data.

4. Comparison of network forensics techniques

4.1 Mechanism: The investigation process of various mechanisms is based on the information of network logs, network packets, and various network events of the network.

- Logging (LO): Used to record network flows and patterns in database to determine evidences regarding attacks. However, it faces challenges in terms of less storage capacity to store all network flows, protecting security devices and fast computation at the point of huge network traffic flows. *Some mechanisms that use network logs are random moonwalk algorithm, Apriori algorithm, hypothesis generation, immune approach, steganography, and pattern and protocol analysis.*
- Packet Marking (PM): Mark network packets at different routers during network flow from sender to its destination and is used by IP traceback techniques to identify sender IP address that is spoofed by intruders. However, its becomes problematic when intruder sends huge amount of packets and because of routers low memory. There 3 types of packet marking techniques: deterministic, packet marking at every router, iTrace. *Some mechanisms that use PM are Authenticated Evidence Marking Scheme (AEMS), tree analysis algorithm, Probabilistic Packet Marking (PPM).*
- Heuristic Based (HB): Used to observe and solve the problems based on the network information. *Some mechanisms that use HB are Spread spectrum technique, immune theory, attack graphs, finite state machine, Hidden Markov model, and fuzzy cognitive maps.*

4. Comparison of network forensics techniques

4.2 Target dataset: Shows the type of data which is targeted by NFT. For instance, data traffic, converge network such as VoIP data, and network model.

4.3 Target instance: Represents the type of instance which is targeted by NFT to identify digital evidence. These instances include complete packet, packet header, and network nodes.

4.4 Forensic processing (FP): Depicts the way network forensics takes place according to its location, such as centralized or decentralized.

- Centralized forensic: A single forensic server is accessed by different network nodes or agents locally or remotely, with quick response time and less time delays but with lack of scalability, more focus by attackers and single point of failure.
- Decentralized forensic: A distributing forensic server in the network is used, but with many limitations such as less consistence, time delays, lack of centralized control, difficult synchronization among distributed data, complexity, higher overhead, use more resources, and high bandwidth communication channels are required.

4. Comparison of network forensics techniques

4.5 Time of investigation: Illustrates either network forensics is performed statically or dynamically.

- Static forensics: the investigation is performed after the attack, so it traces out each and every event properly from the network logs and trace out intruders activities briefly and accurately, but there is a risk of overwriting existing data due to lack of storage capacity and there is no guarantee the data is not altered by the intruder.
- Dynamic forensics: Also called live forensics, where network data is captured, recorded and analyzed at the time of its flow, so it is useful for large distributed networks, but it requires more computational resources and a large amount of data storage space.

4. Comparison of network forensics techniques

4.6 Execution definition: Refers to the type of approach used for investigation. These approaches are divided in proactive and reactive.

- Proactive: Used to investigate the incident in real-time by providing automation to the system while minimizing user intervention. It provides more reliable and accurate evidence in real-time, provides early detection of network attacks and reduces the chance of deleting evidence by intruders after the attack. However, it increases processing and storage overheads in terms of identifying attack patterns and preserving evidence in real-time.
- Reactive: It is a post mortem approach to investigate an attack after it has occurred. Investigates network vulnerabilities by identifying, preserving, collecting and analyzing digital evidence extracted from the network in order to determine root cause of the attack, correlate intruder to the attack, minimize effect of the attack and investigate the malicious incident with reduced processing. However, it is more time consuming and attackers may use anti-forensics techniques to delete traces.

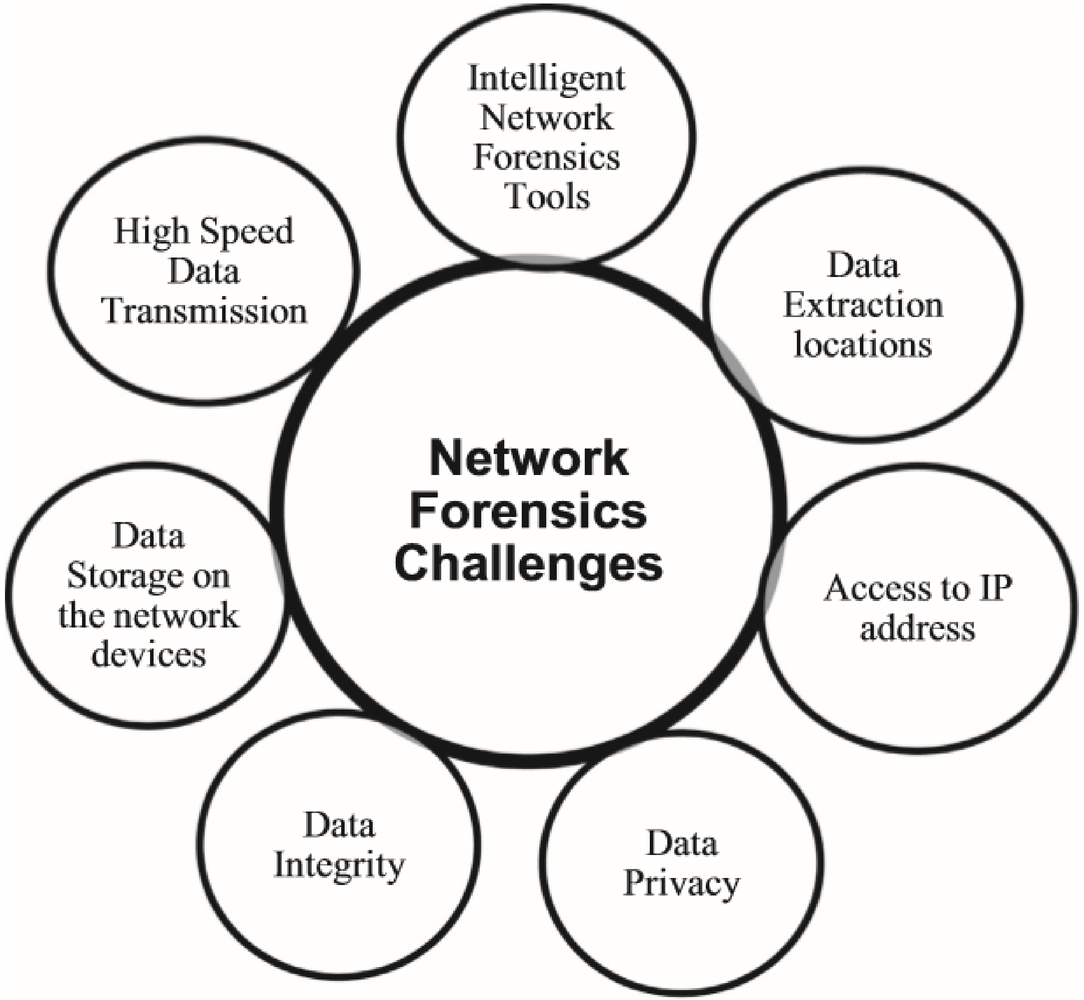
4.7 Objective function: Shows the purpose of proposing a NFT. Different objectives of the network forensics includes origin of attack, visualizing the attacks, reconstructing the attacks, forensics explanation, dynamic forensics, reliability of evidence, analyzing intrusion data, scalable and impact analysis, identifying multi-stage network attacks, evidence collection, identifying worst attacks, event classification, evidence reduction, signature recognition, prevention of novel attacks, and effective feature selection.

4. Comparison of network forensics techniques

Frameworks		Mechanisms			TD	TI	FP	ToI	ED	OB
		LO	PM	HB						
Trace back	NFEA (Kim and Kim, 2011)	No	Yes	No	Data traffic	Packet header	Centralized	Static	Proactive	Origin of attack
	LWIP (Fen et al., 2012a)	No	Yes	No	Data traffic	Packet header	Centralized	Dynamic	Proactive	Origin of attack
	Scalable NF (Chen et al., 2013a)	Yes	No	Yes	Data traffic	Complete packet	Centralized	Dynamic	Proactive	Origin of attack
	HB-SST (Yu et al., 2013)	No	No	Yes	Data traffic	Complete packet	Decentralized	Dynamic	Proactive	Origin of attack
	ITP (Jeong and Lee, 2013)	Yes	No	No	Data traffic	Packet header	Centralized	Static	Reactive	Origin of attack
Converge network	PBNF (Pelaez and Fernandez, 2009)	Yes	No	No	VoIP	Packet header	Centralized	Dynamic	Reactive	Real-time analysis
	VoIP-NFDE (Lin et al., 2010)	Yes	No	No	VoIP	Packet header	Centralized	Dynamic	Reactive	Filtering network traffic
	VoIPEM (Ibrahim et al., 2012)	Yes	No	Yes	VoIP	Complete packet	Centralized	Static	Reactive	Reconstruction of attacks
Attack graph	SA (Albanese et al., 2011)	No	No	Yes	Network model	Network nodes	Centralized	Static	Reactive	Scalable and impact analysis
	MLL-AT (Fen et al., 2012b)	No	No	Yes	Network model	Network nodes	Centralized	Dynamic	Reactive	Identify multi-stage n/w attack
	AGFE (Liu et al., 2012)	No	No	Yes	Network model	Network nodes	Centralized	Dynamic	Reactive	Evidence collection
	FCM (Diamah et al., 2012)	No	No	Yes	Network model	Network nodes	Centralized	Static	Reactive	Identify worst attack
	CSBH (Zhang et al., 2012)	No	No	Yes	Network model	Network nodes	Centralized	Static	Reactive	Cost-benefit security harden
	AGVI (Harbort et al., 2011)	No	No	Yes	Network model	Network nodes	Centralized	Static	Reactive	Visualization
Distribution	ForNET (Shanmugasundaram et al., 2003)	Yes	No	No	Data traffic	Packet header	Decentralized	Static	Proactive	Investigation
	DRNIFS (Ren and Jin, 2005)	Yes	Yes	No	Data traffic	Complete packet	Decentralized	Dynamic	Reactive	Emergence response
	DCNFM (Ren, 2004)	Yes	No	No	Data traffic	Complete packet	Decentralized	Static	Proactive	Origin of attack
	DNF-IA (Wang et al., 2007)	Yes	No	Yes	Data traffic	Packet header	Decentralized	Dynamic	Reactive	Evidence collection
Intrusion detection system	AIDF (Sy, 2009)	No	No	Yes	Data traffic	Complete packet	Decentralized	Static	Reactive	Forensic explanation
	DFITM (Chen et al., 2009)	No	No	Yes	Data traffic	Complete packet	Centralized	Dynamic	Reactive	Dynamic forensics
	IIFDH (Fan and Wang, 2010)	Yes	No	No	Data traffic	Complete packet	Centralized	Dynamic	Proactive	Reliability of evidence
	NFIDA (Jiang et al., 2012)	Yes	No	No	Data traffic	Packet header	Centralized	Static	Reactive	Analyze network intrusion data

LO: Logging; **PM**: Packet marking; **HB**: Heuristic based; **ED**: Execution definition; **TD**: Target datasets;
TI: Target instance; **FP**: Forensic processing; **ToI**: Time of investigation; **OB**: Objective function

5. Open Challenges in network forensics



5. Open Challenges in network forensics

Network forensics challenges	Proposed solutions	Explanation	
Network speed	Specialized hardware e.g. NIFIC	– NIFIC: it contains of gigabit ethernet ports that capture high speed data packets, classify and filter, forward to stated interface, and perform packet analysis in FPGA programmable processing element.	Tripathi (2009)
	Software solution e.g. nCap library	1. Capture packet between 1 and 10 gigabits speed	Deri (2005)
		2. Ability to develop from user space	
		3. Use for active and passive monitoring of the network.	
	Distributed packet capturing	1. Capture packets with load balancing among several nodes	Morariu and Stiller (2008)
		2. Cost effective due to no dedicated hardware required	
		3. Used simultaneously with other packets capturing tools	
Storage capacity	Traffic archiving system with flow granularity	– TIFAFlow: Time machine based packet capturing, perform fast bit indexing and further store it on hard drive. It also increase flow query operation.	Chen et al. (2013b)
	Compress bitmap index in real time on GPU	1. Store up to 185 million records per second	Fusco et al. (2013)
		2. Indexing offloaded to GPU architecture	
		3. CPU intervention is scare	
	Packet-to-disk application: n2disk	1. Capture packet of any size in 10 gigabit at line rate on commodity hardware	Deri et al. (2013)
		2. Can be used for single thread and multi-thread packet consumers	
		3. Configurable to use in real time situation to index packet	
Data integrity	Systematic analysis using GUI-based monitoring	1. Packets are judge by ensuring real time properties. This is performed by collecting servers, which further distribute analyzed information to the clients while also storing it in database	Si-Young and Jong-Chan (2012)
		2. Performs hash function	

5. Open Challenges in network forensics

Data privacy	Forensic attribution	1. It helps investigator to view data of interest through forensic attribution 2. Each observer will verify packet signature whereas it enforce attribution property 3. Aforementioned can be achieved by using following methods (a) Group signatures, (b) BBS short group signatures	Afanasyev et al. (2011)
IP address problems	Source address validation	1. Based on SAVI proposal 2. Binds source host IP, Mac addresses and uplink port properties in layer switches 3. No node can spoof IP addresses of attached node to same uplink	(Bi et al., 2013)
Data extraction locations	Central log repository	– Allow all network traffic to pass through central device installed for monitoring and analyzing.	Didier Stevens (2012)
	Targeting primary network devices	– This may be useful in single event of interest. But this might not provide complete evidence	
Intelligent network forensic tools	Fidelis XPS	1. Capture, visualize, and record session of interest 2. Automatic response, reduce cost, increase bandwidth, and provide proactive awareness 3. Real time visualization	Savchuk (2013)
	WildPackets network forensic tools	1. Capture, record, and analyze in 10 gigabit network traffic speed 2. Analyze data at point of capture in real time situation 3. Comprehensive data collection	McCreery (2012)

6. Conclusion and future directions

- The forensic investigation aims at the origin of the attack, reliability and integrity of the evidence, visualization of attack paths, and determining worst attack paths, which are achievable whenever investigators are clear about the network infrastructure and attack behavior by having appropriate network forensic tools and extensive network forensics knowledge.
- NFT play a vital role in identifying, capturing, recording, and analyzing legal evidences in distributed networks; so they are required to be scalable with increasing network infrastructure in order to analyze fast moving and huge amount of network packets collected at various locations in the network.
- A comprehensive solution is desired in deploying, managing, and bearing less cost for network forensic strategies in distributed networks, resulting in improving and managing easily network security and its visibility in network complexity.
- The development of intelligent network forensic tools to focus on specific type of network traffic analysis is a challenge in terms of future perspective.
- Network forensics at distributed networks of the cloud computing and mobile cloud computing needs to be explored.

Thanks for your attention