Cyberspace Forensics Readiness and Security Awareness Model

Aadil Al-Mahrouqi
School of Computer Science and Informatics
University College Dublin
Dublin, Ireland

Sameh Abdalla
School of Computer Science and Informatics
University College Dublin
Dublin, Ireland

Tahar Kechadi
School of Computer Science and Informatics
University College Dublin
Dublin, Ireland

***Abstract*—The goal of reaching a high level of security in wire- less and wired communication networks is continuously proving difficult to achieve. The speed at which both keepers and violators of secure networks are evolving is relatively close. Nowadays, network infrastructures contain a large number of event logs captured by Firewalls and Domain Controllers (DCs). However, these logs are increasingly becoming an obstacle for network administrators in analyzing networks for malicious activities. Forensic investigators mission to detect malicious activities and reconstruct incident scenarios is extremely complex considering the number, as well as the quality of these event logs. This paper presents the building blocks for a model for automated network readiness and awareness. The idea for this model is to utilize the current network security outputs to construct forensically comprehensive evidence. The proposed model covers the three vital phases of the cybercrime management chain, which are:**

1) Forensics Readiness, 2) Active Forensics, and 3) Forensics Awareness.

Keywords*—Network Forensics; Forensics Readiness; Network Security; Active Forensics; Reactive Forensics; Forensics Awareness and Network Security model

I. INTRODUCTION

The cybercrime landscape has increased dramatically with the use of more sophisticated techniques and greater knowledge of cybercrime. There are many challenges faced by todays digital forensics. The lack of both funding and qualified professionals, as well as cross-jurisdictional legal struggles are just a sample of the main body of issues [1].The first Digital Forensics Research Workshop (DFRWS) [2] was held in Utica, N.Y., in 2001. DFRWS provided the first proper framework and presented guidelines for conducting a technical digital investigation.

It is now evident that the cyber-infrastructure requirements and associated data management systems are becoming large in number, highly dynamic in nature, and exceptionally attractive for cyber-crime activities [3]. Protecting the sensitive data cyber-infrastructure portals are relying on information security for daily activities which is not a trivial task. The techniques used to perform cybercrimes are becoming relatively sophisticated with the firewalls protecting them. Reaching high-levels of data protection in both wired and wireless networks, in order to face recent cybercrime approaches is a challenge that is continuously proving hard to achieve.

Since that first workshop, many scholars have worked to make digital evidence easier to demonstrate by establishing many types of graphs in order to represent evidence and attack scenarios. The scholars utilized a mathematical formula and algorithms to construct these graphs to recognize the patterns of the attack [4]. Unfortunately, most of these graphs provide a high-level, abstract view of the complex attack [5]. Examples of investigation graphs consist primarily of scenario graphs, forensics graphs, logic exploitation graphs, attack graphs, and evidence graphs [6]. The digital systems can be described mathematically as a finite state machine and can represents this information in the form of a graph (nodes and arrows) [7].

Figure 1 shows the cyber-crime management chain, it consists of four stages namely; proactive (readiness), active, reactive and awareness. The first phase in the cyber-crime chain is the proactive phase and its goal is to prepare the target network to automatically prevent and detect the attack or illegal activities before the network gets infected, such as user authentication and system capable of avoiding programming errors and information protection e.g. Privacy Preserving Data Mining (PPDM). The active approach in the cyber-crime chain is used to detect and analyse anomaly activities and attack in real-time e.g. Firewalls. The reactive approach deals with the analysis of the victim network or assesses the incident after it happens e.g. Host-based (HIDs) and Network-based (NIDs) Intrusion detection system. Finally, the awareness approach deals with the training and awareness proposal. These works take into consideration the important factors during forensics investigation, for example; cost, time, low incident impacts, facilities network investigation procedures, high quality outcomes, organization reputation and business activities disruption. Furthermore, the aim to propose an attack and evidence integration graph is to increase the efficiency of investigation results. In addition, the data flow in the proposed model is designed based on the network OSI model. On the other hand, this paper presents a forensics awareness model designed to generate a best practice for system administrations and forensics investigators to learn about security vulnerabilities from previous cases in the network infrastructure, as well as different sources.

The remaining part of this paper is organized as follows; Section 2 outlines previous work. Section 3 describes the proposal model. Section 4 establishes a case study with the aims to give an idea of how to create a criminal graph. Finally,

section 5 deals with the conclusion and some perspectives on future work.

II. Previous Work

The authors [8] performed an in-depth survey for events admissibility in the Irish court of law. Overall, the legal review is mainly focused on different primary areas: the admissibility and authentication of digital evidence and focuses mainly on Irish law. Admissibility refers to a set of lawful tests carried out by a judge for forensic assessment of the finding evidence. Trustworthy means that an accurate copy of digital evidence was acquired, and that it has continued to be unchanged since it was recovered. Authentication is a process to check the reliability of digital evidence. The judge summarizes five issues that must be considered when evaluating whether evidence will be admitted, namely; not unduly prejudicial, best evidence, not hearsay or admissible hearsay, authenticity and relevance.

Wang & Daniels [9], in their proposed evidence graph model seek to facilitate the presentation and manipulation of intrusion evidence. This model aims to reduce the redundancy in firewall output intrusion alerts. The proposed architecture facilitates the evidence presentation process and provides automated intrusion evidence analysis. The evidence module is considered the most important module in the Wang & Daniels proposed architecture because it plays an important role in analysis visualization of capture evidence.

Later, Wang & Daniels [10] proposed diffusion and graph spectral methods. These proposed methods aimed to establish a systematic forensics investigation process framework. Moreover, through these proposals Wang & Daniels attempted to provide high-performance computation methods to be used in the forensics analysis field as a form of well-utilized mathematical science.

In 2004, Gladyshev [11] proposed a formalized approach for Event Reconstruction. This approach was based on the terms of the finite state machine model of computation. The finite state machine model was used to define all possible attack scenarios in the computer network incidents. Furthermore, Gladyshev defined Event Reconstruction 'as a process of finding all potential computations of the machine that agree with the digital evidence of the incidentl;'. The scholar proposed an algorithm for the Event Reconstruction process that consists of three phases. The first phase calls for obtaining the finite state model of the computer system that is under the forensics investigation. In the second phase, all potential attack scenarios of the computer system incidents are defined by using the back trace method from the point in which the cybercriminal was discovered. The third phase calls for rejecting attack scenarios that conflict with the obtainable evidence [12].

Liu & Wijesekera [13] proposed merging sub-evidence graphs with an integrated evidence graph for network forensics analysis. This paper shows how to integrate different evidence graphs with or without the help of a corresponding attack graph. The proposal model assumes that an integrated evidence graph shows all attacks using global reasoning. Consequently, the research provided two algorithms that help integrate evidence graphs with a probabilistic evidence graph.

Fig. 1. Cybercrime Management Chain

Phillips & Swiler [14] proposed an approach for network risk analysis based on an attack graph that defines the set of attack paths that have a high probability of success for the attacker. This approach requires a predefined data-set as input information before starting to use the system. As a result, the system will generate an attack graph based on predefined information.

Sheyner et al. [15] proposed automated techniques in order to establish and generate the attack graphs. The techniques are based on a set of algorithms that are used to reconstruct attack scenarios automatically. After that, the reconstructed attack scenario is represented in the attack graphs. The visual representation of attack graphs allows forensics investigators to easily understand the attack scenario in an efficient manner. The authors implemented a network forensics tool based on the proposed algorithms, testing it in a small Local Area Network (LAN) that consists of an intrusion detection system and firewalls.

Bruaschi et al. [16] proposed a model that can organize digital forensics knowledge in a reusable way. In other words, this model can reuse the gathering techniques and some hypotheses in order to find the best guideline for hypotheses formulation.

III. Cyberspace Forensics Readiness and Security Awareness Model

In figure 2 shows the overall view of the proposed cyberspace forensic readiness and security awareness model.

Logs classification processes submodel: Basically, the op- erating system in the network firewalls and domain controllers (DCs) are able to classify the computer network and system events logs into predefined groups. This model was designed to increase the filtering process of the output events logs. It will classify the output logs into different groups, namely alerts and information.

Alert logs collection model: This model is designed to collect only the alert logs. These logs will be stored in the alert logs warehouse.

Alerts prepossessing model: The stored alert logs contain redundancy data and irrelevant information [11]. The alerts preprocessing model is used to filter out all redundancy data and irrelevant information from the alert logs. The alerts preprocessing model has two stages; format standardization and redundancy management. The format standardization process aims to convert the different event logs formats into one unified, common syntax format while the redundancy management process aims to reduce the duplication of the single event.

Assets Knowledge warehouse: Assets knowledge warehouse is designed to store basic information of all assets available in the network infrastructures.

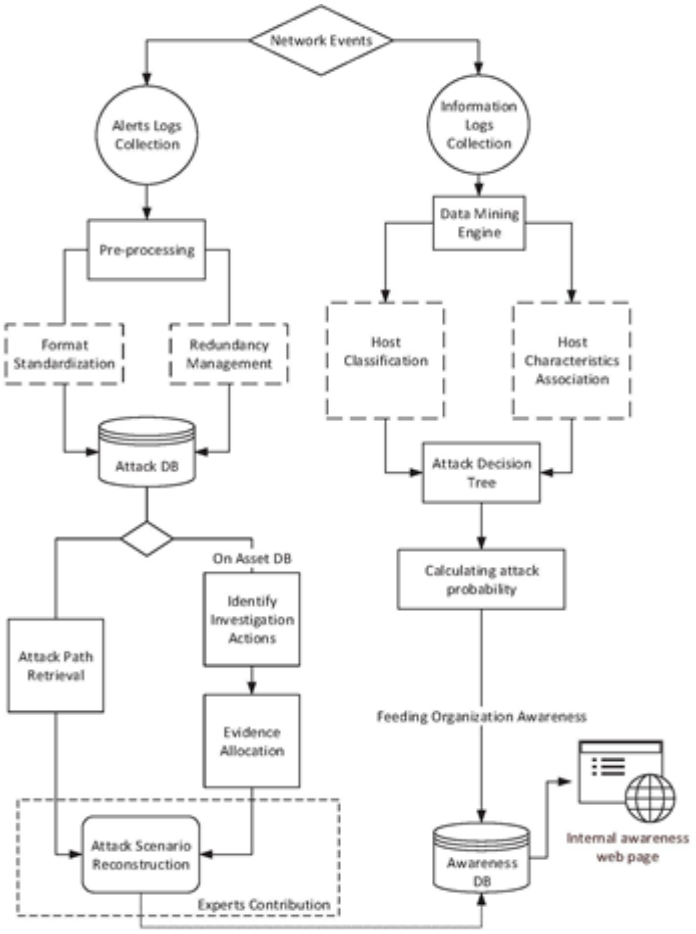

Fig. 2. Network Forensics Readiness and Security Awareness Framework

Attack knowledge warehouse: The assets knowledge warehouse is designed to store basic information of all assets available in the network infrastructures.

Attack path Retrieval: An attack graph provides a visual representation of the attack paths as well as evidence for each node (host) in each path (see figure 4). The attack paths describe all exploited network assets. The attack graphs will be generated based on databases, namely asset knowledge and attack knowledge. The nodes indicate the exploited hosts while the edges indicate the security vulnerability used to hack the host. The information shown in this graph is based on a chain of custody manner.

Scenario reconstruction submodel: After generating the attack and evidence graphs, this model is used to reconstruct the attack scenario. This process will reprocess the criminal graph with the help of criminology sciences and hypothesis expert knowledge.

A. Information prepossessing model

Information collection sub model: The information collection submodel will collect all output of information logs from event log classification processes and forward it to the information logs warehouse.

TABLE I. THE EVENTS STRUCTURE

The Events structure field	
Field	Description
Type	Shows the type of events (Information, Warning, Error, etc.)
Time	Shows the time of the event happened
Date	Shows the date of the event happened
Event ID	Shows an event log number that identifies the event type
Device	Shows the device where the event happened
User	Shows the computer user who has generate the events or who logged to the computer system when the happened
Source	Shows the source produced the event

Data mining Engine: As there are so many information logs in the information logs warehouse, it is very difficult to check all of them and update information security awareness. This step is used to convert information logs into an easier format that will be useful for security information awareness.

The data mining engine consists of two types of processes; host classification types and host characteristics associations. First, the host classification process will be used to classify all existing assets in the network infrastructure into certain groups based on host types, such as router, switches, domains controllers, firewalls, etc. Second, host characteristics associations will be used to associate each log to the appropriate predefined group. Using the association process, it will be necessary to analyze the logs header format to be able to know the appropriate predefined group.

Calculating attack probability: Calculating the attack probability process will be used to process the output results of the attack decision tree. This process examines the attack probability for each asset (for example, the file server probability affected by DOS) based on previous experiments knowledge through the data-set analysis.

Awareness DB: The awareness DB is used to store the attack probability for each asset in the network infrastructure. This database feeds the internal awareness Web page through security awareness and vulnerabilities for network assets.

B. The normalization process of alerts and information logs

As mentioned earlier, the normalization process will be used to convert the event logs formats into a unified format. This process will help to aggregate the logs and reduce redundancy and noise information. Table I I shows a proposed unified structure field of event logs.

C. The relationship between the evidence and attack

It is very important to know the relationship between the evidence and attacks. This relationship helps us in the investigation process, as well as increases the admissibility of the investigation case in court. Moreover, the increased amount of evidence related to a specific attack case will increase the background information about the attacker. There are different types of relationships between the detected evidence and the attacks, including one-to-one, one-to-many, and many-to-one.

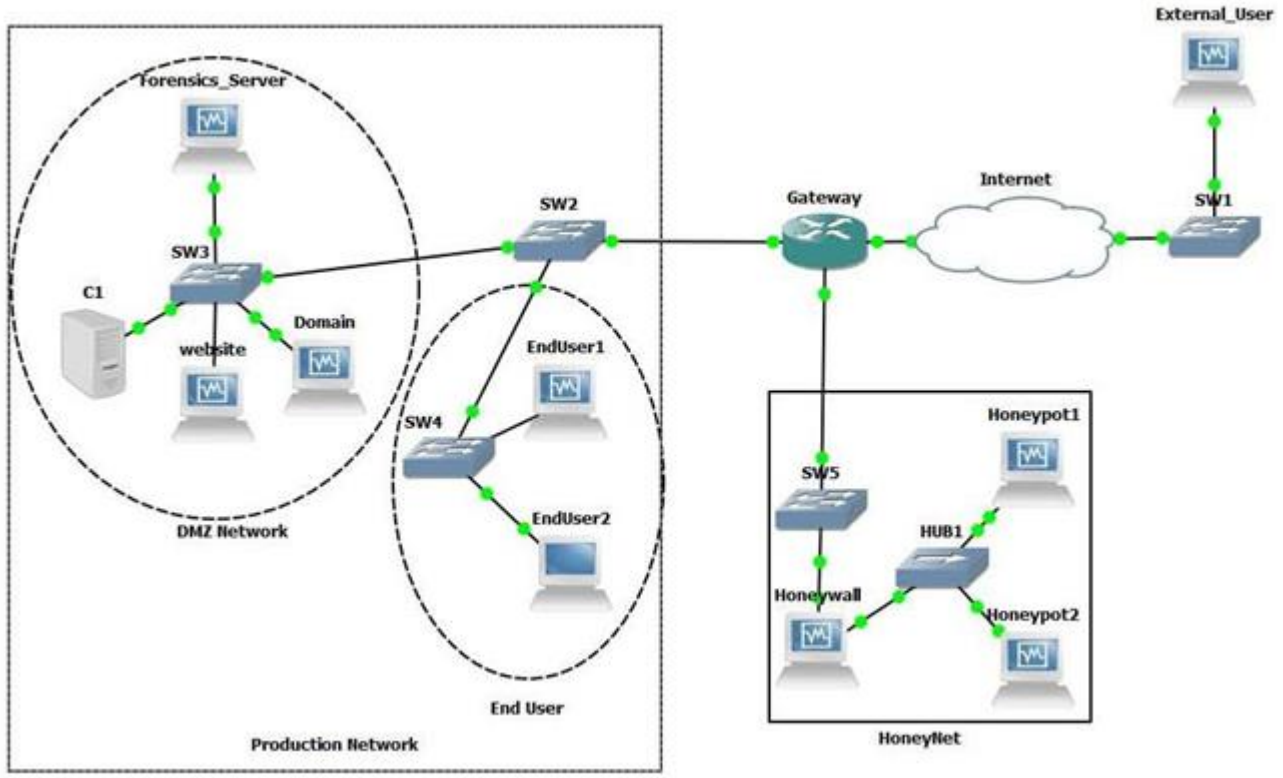

Fig. 3. Simulation Honeynet Network in GNS3

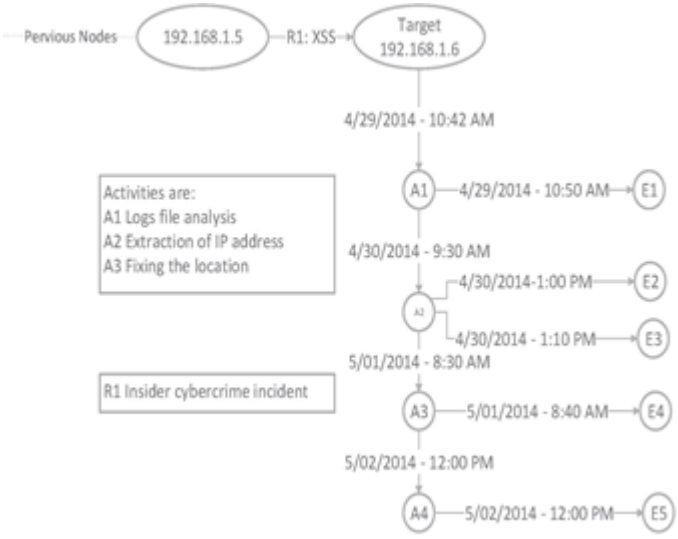

Fig. 4. Integration between Attack Graph and Evidence Graph

IV. Example of Criminal Graph

This paper presents a picture of the proposed graph that integrates the attack and evidence graphs. Section IV-A establishes a criminal scenario.

A. Simulating SQL-Injection Cyber-attack using GNS3

The authors [17] presented a simulation study network attack scenario. This is the first step towards validating the proposed model. The simulation case study used capturing, normalizing and analyzing events that are introduced in section III. The main point of designing virtual network attack environments is to create a sandbox that allows one to perform such experiments, from real assets and at a low cost. Both the capturing and examination of the events were conducted in the simulated case study. The detection of network artifices changes after the execution of SQL-Injection attacks were also recorded. The outcome of this experiment can be used as a recommendation in real cyber-infrastructure. The core idea of the case study is to examine the website that has been compromised by an SQL injection attack. To simulate this attack scenario many open source tools were used such as Graphical Network Simulator (GNS3), Oracle VM Virtual Box and VMWare workstation. The wireshark forensics tool was also used to detect criminal activity from the network layer (Layer 3 in OSI model), in addition, the victims and attackers devices by using the Volatility Framework 2.4 were also examined.

Simulation approaches helps to graphically simulate an attack for courts, Jury and Investigators. The simulation approaches also helps to simplify the incidence (1 Image = 1000 words). The current study [17] proposes Investigation learning methodology based on the proposed case study. The learning methodology consists of two stages; stage one is to build a network topology of the proposed case study and stage two is to create a network union Matrix.

This approach allows specific network devices configuration to be simulated, perform SQL injection attacks against victim machines and collect network logs. The main motivation of thiswork is to finally define an attack pathway prediction methodology that makes it possible to examine the network artifacts collected in case network attacks.

Figure 4 shows the integration between the attack graph and the evidence graph. Moreover, nodes indicate compromised

hosts while edges refer to security vulnerabilities used by the attacker. Moreover, under each compromised host, there is another graph that shows the series of actions carried out by the forensics investigators. Furthermore, the actions carried out by the forensics investigation are linked to another graph called the evidence graph. This graph will show the output evidence as a result of each forensics investigative action.

The authors purposed a new network forensics model [8] that can makes network events admissible in the court of law. The present model collects available logs from connected network devices, applies decision tree algorithm in order to filter anomaly intrusion, then re-route the logs to a central repository where events management functions are applied.

V. Conclusions and Future Work

This proposed model contains approximately fifteen different models. The proposed models work as a single unit in order to process and normalize the captured network logs. The main point of designing the model is to find a way to forensically visualize the evidence and attack scenario in a computer system. Moreover, this paper listed some methods and approaches proposed by scholars to construct the attack scenario. Nevertheless, the graph representation is one of the best approaches used in the forensics investigation; the researchers in this field have proposed several types of graphs, including scenario graphs, logic exploitation graphs, forensics graphs, attack graphs, and evidence graphs.

Since the attempt to reconstruct scenarios of network attacks from collected data (i.e., alarms, alerts and logs) requires brain-like reasoning to understand these events. Therefore, Bio-inspired approaches [18] to self-organizing network events and creating the linkage between them are of relevance to the research studies. The future plans is to examine the possibility to replace the traditional database approach to storing events with a bio-inspired mechanism and, study the affect of that on the quality of the scenarios produced.

This model acts as a first step toward network logs analysis. The future work will focus on involving mathematics and algorithm science for each proposed blocks to help validate the model. Furthermore, trying to utilize criminology science to enhance any future proposed models or approaches are a key priority.

Acknowledgment

The authors would like to acknowledge the support of by the Insight Centre for Data Analytics and UCD Centre for Cybersecurity and Cybercrime Investigation.

References

[1] I. Baggili and M. Kiley, “Digital forensics a brief overview of critical issues,” Digital Forensics Investigator News, 2008.

[2] G. Palmer, “A road map for digital forensics research-report from the first digital forensics research workshop (dfrws),” Utica, New York, 2001.

[3] A. Salim Al-mahrouqi, S. Abdalla, and T. Kechadi, “E-government alerts correlation model,” in Qatar Foundation Annual Research Conference, no. 1, 2014, p. ITPP1120.

[4] V. Leucari. (2005) Analysis of complex patterns of evidence in legal cases: Wigmore charts vs. bayesian networks. [Online]. Available: https://www.ucl.ac.uk/jdi/research/evidence-network/docs/BURGLARY.PDF

[5] L. P. Swiler, C. Phillips, D. Ellis, and S. Chakerian, “Computer-attack graph generation tool,” in DARPA Information Survivability Conference & Exposition II, 2001. DISCEX’01. Proceedings, vol. 2. IEEE, 2001, pp. 307–321.

[6] S. Neralla, D. L. Bhaskari, and P. Avadhani, “A novel graph model for e-mail forensics: Evidence activity analysis graph,” International Journal of Engineering Science and Technology, vol. 5, no. 10, p. 1750, 2013.

[7] J. James, P. Gladyshev, M. T. Abdullah, and Y. Zhu, “Analysis of evidence using formal event reconstruction,” in Digital Forensics and Cyber Crime. Springer, 2010, pp. 85–98.

[8] A. Al-Mahrouqi, S. Abdalla, and T. Kechadi, “Efficiency of network event logs as admissible digital evidence,” in Science and Information Conference 2015, London, United Kingdom, 28-30 July 2015, 2015.

[9] W. Wang and T. E. Daniels, “Building evidence graphs for network forensics analysis,” in Computer Security Applications Conference, 21st Annual. IEEE, 2005, pp. 11–pp.

[10] ——, “Diffusion and graph spectral methods for network forensic analysis,” in Proceedings of the 2006 workshop on New security paradigms. ACM, 2006, pp. 99–106.

[11] P. Gladyshev, “Formalising event reconstruction in digital investigations,” Ph.D. dissertation, University College Dublin, 2004.

[12] M. Sebastian and P. Chandran, “Towards designing a tool for event reconstruction using gladyshev approach,” in Proceedings of the 2011 ACM Symposium on Applied Computing. ACM, 2011, pp. 193–194.

[13] A. C. Liu and D. Wijesekera, “Merging sub evidence graphs to an integrated evidence graph for network forensics analysis,” Advances in Digital Forensics IX, pp. 227–241, 2013.

[14] C. Phillips and L. P. Swiler, “A graph-based system for network-vulnerability analysis,” in Proceedings of the 1998 workshop on New security paradigms. ACM, 1998, pp. 71–79.

[15] O. Sheyner, J. Haines, S. Jha, R. Lippmann, and J. M. Wing, “Automated generation and analysis of attack graphs,” in Security and privacy, 2002. Proceedings. 2002 IEEE Symposium on. IEEE, 2002, pp. 273–284.

[16] D. Bruschi, M. Monga, and L. Martignoni, “How to reuse knowledge about forensic investigations,” in Digital Forensics Research Workshop, 2004, pp. 10–13.

[17] A. Al-Mahrouqi, P. Tobin, S. Abdalla, and T. Kechadi, “Simulating sql-injection cyber-attacks using gns3,” International Journal of Computer Theory and Engineeringl, vol. 8, no. 3, pp. 213–217, 2015.

[18] D. Floreano and C. Mattiussi, Bio-inspired artificial intelligence: theories, methods, and technologies. MIT press, 2008.