网络安全与网络取证：机器学习方法

Ibrahim Goni[1], Jerome Mishion Gumpy[2], Timothy Umar Maigari[3], Murtala Muhammad[4], Abdulrahman Saidu[4]
1 尼日利亚 穆比 阿达马瓦州立大学计算机科学系
2 尼日利亚 加舒阿 联邦大学计算机科学系
3 尼日利亚 贡贝 联邦教育学院计算机科学系
4 尼日利亚 塔拉巴 巴厘联邦理工学院计算机科学系

摘　要：云计算和物联网的发展导致了世界各国（发达国家和发展中国家）的互联互通，全球网络为互联互通提供了平台。数字取证是一个计算机安全领域，它使用软件应用程序和标准指南，支持从任何计算机设备中提取证据，这些设备完全足以让法庭根据所获得信息的全面性、真实性和客观性进行使用和判断。由于物联网中每天都在发生的攻击、威胁、病毒、入侵等最新形式，网络安全是全世界互联网用户最关心的问题。这项工作的目的是对机器学习算法在网络安全和网络取证中的应用进行系统回顾，根据这一发现，对机器学习方法在网络取证和网络安全中的最新应用进行了系统调查，还指出，网络安全有十个步骤；网络安全、用户教育和意识、恶意软件预防、可移动媒体控制、安全配置、管理用户权限、事件管理、监控以及家庭和移动工作，为深入学习、计算智能、软计算在网络安全和网络取证中的应用的进一步研究方向铺平道路。
关键词：网络安全；网络取证；网络空间；网络威胁；机器学习和深度学习

Cybersecurity and Cyber Forensics: Machine Learning Approach

Ibrahim Goni[1], Jerome Mishion Gumpy[2], Timothy Umar Maigari[3], Murtala Muhammad[4], Abdulrahman Saidu[4]
1.Department of Computer Science, Adamawa State University, Mubi, Nigeria
2.Department of Computer Science, Federal University, Gashua, Nigeria
3.Department of Computer Science, Federal College of Education Gombe, Nigeria
4.Department of Computer Science, Federal Polytechnic Bali, Taraba Nigeria

Abstract: The proliferation of cloud computing and internet of things has led to the connectivity of states and nations (developed and developing countries) worldwide in which global network provide platform for the connection. Digital forensics is a field of computer security that uses software applications and standard guidelines which support the extraction of evidences from any computer appliances which are perfectly enough for the court of law to use and make a judgment based on the comprehensiveness, authenticity and objectivity of the information obtained. Cybersecurity is of major concerned to the internet users worldwide due to the recent form of attacks, threat, viruses, intrusion among others going on every day among internet of things. The aim of this work is make a systematic review on the application of machine learning algorithms to cybersecurity and cyber forensics, systematic survey method was used on recent application of machine learning algorithms on cyber forensics and cyber security based on this findings it is observed that cybersecurity is based on confidentiality, integrity and validity of data, it is also noted that there are ten steps to cybersecurity; network security, user education and awareness, malware prevention, removable media control, secure configuration, managing user privileges, incident management, monitoring and home and mobile working and pave away for further research directions on the application of deep learning, computational intelligence, soft computing to cybersecurity and cyber forensics.
Keywords: Cybersecurity; Cyber forensics; Cyber space; Cyber threat; Machine learning and deep learning

1. 引言

网络空间为当今的技术提供了居住环境和平台，从物联网、5G、Fog、边缘等，它局限于发展和扩展，并支持各种科技创新，但有好有坏[1，3]。根据 2017 年全球网络安全指数显示，世界人口中近一半（35 亿用户）连接到网络空间，他们进一步估计，到 2020 年，网络空间将有 120 亿设备到设备连接。另据报道，到 2020 年，地球上 80% 的成年人将拥有智能手机[48]，此外，49.7%

的总人口将连接到互联网，与 2000–2017 年相比，全球增长了 936%[49]。然而，对这一空间的威胁日益令人担忧。这项研究工作的目的是探索机器学习算法在网络安全和网络取证中应用的关键研究贡献。

2. 网络取证

数字取证或计算机取证是法医学的一个分支，描述了对发生在计算机网络或计算机系统中的犯罪进行取证调查的技术，这些犯罪被用作网络攻击或进行犯罪活动的武器，但无论使用何种数字设备实施犯罪 [2]。

此外，Nickson 等人 [3] 还将网络取证描述为计算机安全的一个分支，该分支使用软件和预定义技术，旨在从任何形式的数字设备中提取证据，并可提交给法院进行刑事和/或民事诉讼，前提是满足这三个条件：全面性、真实性和客观性。此外，他们能够揭示数字法医报告应该能够显示有关证据的重要事实；比如谁获得了证据？证据在哪里捕获和存储，以及证据发生了什么。 此外，2017 年，Rukayat 等人 [4] 确定了数字取证的目标，即识别证据、记录犯罪、收集和保存证据、包装证据和以不受阻碍的方式运输证据。2017 年，Anwar&Riadi[38] 建议，网络取证依赖于事件的收集和分析，以探索了解和显示绕过安全机制的复杂安全漏洞。在 [46] 中，“数字取证可以说是系统开发中的一个科学框架，用于识别、定位、检索和分析来自计算机、计算机存储介质和其他电子设备的证据，并在法庭案件中呈现发现结果”。数字取证由 Mark[50] 以图形方式表示为。

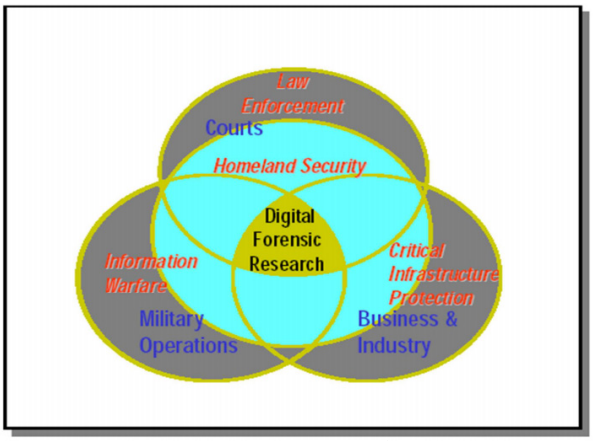

图 1. 数字取证科学 [50]。

2.1 网络犯罪

网络威胁、攻击和破坏已成为互联网用户日常生活中的常见事件 [6]。此外，David 等人 [10] 确定网络恐怖主义是网络度量和恐怖主义的集合体，他们还认为网络恐怖主义是非法使用数字设备进行破坏，不适当地迫使恐吓或进一步恐怖分子的社会经济政治或宗教议程。凯捷研究所透露，在一个案例中，黑客能够访问 27000 辆汽车的 GPS，导致发动机关闭。“迫切需要更多的研究和工具开发，以帮助数字调查人员获取和分析智能手机、平板电脑、可穿戴设备、SatNav 系统、游戏机、汽车、物联网系统和云环境中日益增多的数字证据 [5]。

2.2. 网络安全

网络安全涉及数据安全、网络安全和计算机安全。许多研究人员还将其视为安全防范的应用，以提供数据的机密性、完整性和可用性 [28]，但网络安全的主要目标是预防检测和反应。此外，中央情报局透露，网络安全的主要目标是保密性、完整性和可用性。英国国家网络安全中心列举了网络安全的十个步骤；网络安全、用户教育和意识、恶意软件预防、可移动媒体控制、安全配置、管理用户权限、事件管理、监控以及家庭和移动工作 [35]。此外，Gyun[33] 还透露，人工智能和机器学习是行为建模、零日攻击和高级持续威胁的最重要网络工具。

3. 网络威胁

据美国情报界称，2016 年和 2017 年，乌克兰和沙特阿拉伯遭到了国家赞助的网络攻击，导致政府和非政府组织的主要基础设施遭到攻击。他们进一步指出，已知网络安全威胁被分类为已知网络安全风险，被分类为三个类别；身份盗窃，包括：网络钓鱼、欺骗、伪装、社交工程和密码破解。未经授权的访问包括：目标数据挖掘、后门、窃听和窃听。拒绝服务（DoS、DDoS）包括：逻辑炸弹和密码锁。 网络安全风险投资公司确定，2019 年勒索软件将损失高达 115 亿美元 [44]。“针对英国国家卫生服务的勒索软件攻击影响了 60 个健康信托机构、150 个国家和 20 多万个计算机系统”[45]。

4. 机器学习算法的基本概念

机器学习是一种使用算法分析数据、从数据中学习并基于收集的数据做出决策、预测、检测、分类、模式识别、响应和聚类的技术。这些算法非常依赖于统计和数学优化。在更广泛的意义上，机器学习算法用于聚类、回归、（单变量和多变量）异常检测和模式识别 [34]。机器学习的三种类型是：监督学习、无监督学习和强化。监督学习算法是需要数据集来训练和测试性能的机器学习算法。该数据集必须进行标记，并由定义事件或对象的特征以及预期输出组成。最常见的监督学习算法是决策树、逻辑回归、支持向量机、相关性向量机、随机森林、K NN、bagging 神经网络、线性回归和天真贝叶斯 [33]。无监督学习算法是一种机器学习算法，需要未标记的数据集来训练和测试系统性能。无监督学习中使用的两种主要技术是主成分分析（PCA）和聚类。最常见的无监督学习算法，特别是在安全领域中使用的是分层、k 均值、混合模型、DBSCAN、OPTIC、自组织映射、Bolzan 机器、自动编码器、对抗网络 [34]。

5. 文献综述

2019 年，Bandir[7] 探索了聚类算法的使用，如 K–means 分层聚类、K–means 核、潜在 drichlet 分配和自组织映射技术，用于使用大量数据中的文本聚类进行取证分析。2018 年，Al Jadir 等人，[8] 提出了一种使用

模因算法的鲁棒取证分析方法。2018 年，Sunil&Preeti[9] 揭示了人工智能技术如何应用于网络攻击和安全漏洞。[16] 中，机器学习算法被用于对 android 系统中的恶意软件进行分类。[15] 中，机器学习和深度学习算法被结合并用于网络安全系统。[14，51] 中，机器学习算法也应用于入侵检测系统。[13] 的研究对将机器学习算法和数据挖掘技术结合到网络安全中的研究进行了系统调查。2018 年，Apruzze 等人 [12] 介绍了机器学习和深度学习在网络安全特征中的有效性。在机器学习、深度学习和人工智能技术在网络安全、攻击、入侵检测系统和网络安全中的应用方面，进行了许多调查和系统综述 [16，18-22]。[29] 中，机器学习算法也被用于研究网络安全。[30] 使用模糊逻辑设计了安全框架。

此外，机器学习算法深度学习算法被应用于入侵检测系统，如 A. Abubakar 等人 [23] 的研究所述，基于机器学习的系统可用于检测软件定义网络的入侵。[24] 对基于异常的入侵检测系统进行了广泛的综述。[25] 将机器学习算法应用于异构客户端网络中移动云中的入侵检测。在 [26] 云计算混合入侵检测系统的工作中。[27] 他们使用机器学习算法为工业网络异常检测提供路线图。[31] 提出了汽车网络异常检测系统。深度神经网络和模糊逻辑用于识别网络流量中的异常 [32]。[40] 对用于恶意软件检测的技术进行了系统调查，而 W. Songyang 等人 [41] 使用 API 和机器学习算法来检测安卓中的恶意软件。2015 年，Anastasia 和 Gamayunov[42] 对基于并行和分布式网络的移动设备中的恶意软件检测进行了综述。[43] 他们对静态、动态和混合技术用于恶意软件检测进行了比较分析。还对 WhatsApp 信使进行了取证分析，以识别那些正在使用该应用程序实施犯罪或从事非法业务的人，如 [36-39] 的研究中所述。除了 Parag[47] 之外，还提出了数字取证框架，并与其他没有人工智能技术的框架进行了比较分析。然而，与我们提出的框架相比，该框架没有即时检测和发送信号。

6. 结论

机器学习算法最近被应用于以下网络安全领域，如网络安全、数据安全、端点安全、身份访问安全、云安全、物联网安全、雾安全，但大多数安全系统依赖于检测、预测和响应。此外，网络安全的主要目标是保密性、完整性和可用性。在这项研究工作中，还注意到网络安全有十个步骤；网络安全、用户教育和意识、恶意软件预防、可移动媒体控制、安全配置、管理用户权限、事件管理、监控以及家庭和移动工作。除了人工智能和机器学习之外，它们是行为建模、零日攻击和高级持续威胁的最重要的网络工具。

参考文献

[1] Shahzad S. (2015) protecting the integrity of digital evidence and basic human rights during the process of digital forensics. Ph. D. thesis Stockholm University.

[2] Abdalzim A. M. A. & Amin B. A. M. (2015) a survey on mobile forensics for

android smart phones IOSR Journal of computer engineering 17 (2) 15-19

[3] Nickson M. K., Victor R. K. & Venter H. (2019) Divergency deep learning cognitive computing techniques into cyber forensics Elservier Forensics Science international synergy 1 (2019) 61-67.

[4] Rukayat A. A., Charles O. U. & Florence A. O. (2017) computer forensics guidelines: a requirement for testing cyber crime in Nigeria now?

[5] Casey E. (2016) Editorial- A sea change in digital forensics and incident response. Digital investigation evidence Elservier Ltd 17, A1-A2.

[6] Ehsan S. & Giti J. (2019) Seminars in proactive artificial intelligence for cyber security consulting and research, Systematic cybernetics and informatics 17 (1) 297-305

[7] Bandir A. (2019) Forensics analysis using text clustering in the age of large volume data: a review. International journal of advanced computer and application. 10 (6), 72-76.

[8] Al-Jadir I., Wong K. W., Fing C. C. & Xie H. (2018) Enhensing digital forensics analysis using memetic algorithm feature selection method for document clustering 2018 IEEE international conference on systems, Man and cybernetics 3673-3678.

[9] Sunil B. & Preeti B. (2018) Application of artificial intelligence in cyber security. International journal of engineering research in computer science and engineering 5 (4), 214-219.

[10] David O. A., Goodness O. & Etecte M. A. (2019) Unbated cyber terrorism and huma security in Nigeria. Asian social science 15 (11), 105-115.

[11] April (2014) threat start-SMS spam volume by month of each region SC magazine. available online at http// www.scmagazine.com/april-2014-threat-stats/slideshowz.

[12] Apruzze G., Colajanni M. F., Ferreti L., & Marchett M. (2018) on the effectiveness of machine learning for cyber security in 2018 IEEE international conference on cyber conflict 371-390.

[13] Buckza A. L. & Guven E. (2016) A survey of data mining and machine learning

methods for cyber security intrusion detection IEEE communication survey and totorials 18 (2), 1153-1176.

[14] Biswas S. K. (2018) intrusion detection using machine learning: A comparison study. International Journal of pure and applied mathematics 118 (19), 101-114.

[15] Y. Xin, Kong L., Liu Z., Chen Y., Zhu H., Gao M., Hou H., & Wang C. Machine learning and deep learning methods for cyber security. IEEE Access 6: 35365-35381 (2018).

[16] N. Miloseivic, Denghantanh A., Choo K. K. R. Machine learning aided android malware classification. Computer and electrical engineering 61: 266-274 (2017).

[17] B. Geluvaraj, Stawik P. M., Kumar T. A. the future of cyber security: the major role of Artificial intelligence, Machine learning and deep learning in cyber space. International conference on computer network and communication technologies Springer Singapore. 739-747 (2019).

[18] H. Mohammed B., Vinaykumar R., Soman K. P. A short review on applications of deep learning for cyber security (2018).

[19] M. Rege, Mbah R. B. K. Machine learning for cyber defense and attack. in the 7th International conference on data analysis 73-78 (2018).

[20] D. Ding, Hang Q. L., Xing Y., Ge X., and Zhang X. M. A survey on security control and attack detection for industrial cyber physical system. Neuro-computing. 275. 1674-1683 (2018).

[21] D. Berman S., Buczak A. L., Chavis J. S., Corbelt C. L. A survey of deep learning methods for cyber security information 10 (4): (2018).

[22] Y. Wang, Ye Z., Wan P., Zhao J. A survey of dynamicspectrum allocation based on reinforcement learningalgorithms in cognitive radio network. Artificial intelligence review. 51 (3): 413-506 (2019).

[23] A. Abubakar, Paranggono B. Machine learning based intrusion detection system for software defined networks. 7thInternational conference on Emerging security techniques IEEE. 138-143. (2017).

[24] S. Jose, Malathi D., Reddy B., Jayaseeli D. A survey on anomaly based host intrusion detection system. Journal of physics. Conference series 1000 (1): (2018).

[25] S. Dey, Ye Q., Sampalli S. A Machine learning basedintrusion detection scheme for data fusion in mobile cloud involving heterogeneous clients network. Information fussion 49: 205-215 (2019).

[26] P. Deshpande, Sharma S. C., Peddoju S. K., Junaid S. HIDS: a host based intrusion detection system for cloud computing environment. International journal of system assuarance engineering and management. 9 (3): 567-576 (2018).

[27] M. Nobakht, Sivaraman V., Boreli R. A host-Based Intrusion detection and mitigation framework for smart IoT using open flow in 11th International conference on availability reliability and security IEEE. 147-156 (2016).

[28] A. Meshram, Christian H. Anomaly detection in industrial networks using machine learning: A road map. Machine learning for cyber physical system Springer Berlin Heldelberg. 65-72 (2017).

[29] R. Devakunchari, Souraba, Prakhar M. A study of cyber security using machine learning techniques. International journal of innovative technology and exploring engineering. 8 (7): 183-186 (2019).

[30] E. Alison N. FLUF: fuzzy logic utility framework to support computer network defense decision making IEEE (2016).

[31] A. Taylor, Leblanc S., Japkowicz N. Anomaly detection in auto-mobile control network data with long short term memory network in data science and advance analytics. IEEE international conference. 130-139 (2016).

[32] O. Amosov S., Ivan Y. S., Amosovo S. G. Recognition of abnormal traffic using deep neural networks and fuzzy logic. International Multi-conference on industrial engineering and modern technologies IEEE (2019).

[33] M. Gyun L. Artificial Intelligence for development series: Report on AI and IoT in Security Aspect. (2018).

[34] L. Matt. Rise of machine: machine learning & its cybersecurity applications, NCC group white paper (2017).

[35] National cyber security center UK, www.ncsc.gov.uk.

[36] A. Nuril, Supriyanto (2019) Forensic Authentication of WhatsApp Messenger Using the Information Retrieval Approach. International Journal of Cyber Security and Digital Forensics (IJCSDF) 8 (3): 206-212 (2019).

[37] A Marfianto, I Riadi. WhatsApp Messenger Forensic Analysis Based on Android Using Text Mining Method. International Journal of Cyber Security and Digital Forensics (IJCSDF) 7 (3): 319-327 (2018).

[38] N Anwar, I. Riadi. Forensic Investigative Analysis of WhatsApp Messenger Smartphone Against WhatsApp WebBased, Journal Information Technology Electromagnetic Computing and Information, 3 (1): 1-10 (2017).

[39] S. Ikhsani and C. Hidayanto, Whatsapp and LINE Messenger Forensic Analysis with Strong and Valid Evidence in Indonesia. Tek. ITS, 5 (2): 728-736 (2016).

[40] M. Ashawa, S. Morris. Analysis of Android Malware Detection Techniques: A Systematic Review. International Journal of Cyber Security and Digital Forensics (IJCSDF) 8 (3): 177-187 (2019).

[41] W. Songyang, Wang, P., Zhang, Y. Effective detection of android malware based

on the usage of data flow APIs and machine learning: Information and Software Technology, 75: 17--25 (2016).

[42] Anastasia, S., Gamayunov, D.: Review of the mobile malware detection approaches: Parallel, Distributed and NetworkBased Processing (PDP). In: Proc. 2015. IEEE 23rd Euro micro International Conference, pp. 600—603 (2015).

[43] D. Anusha, Troia, F. D., Visaggio, C. A., Austin, T. H., Stamp, M.: A comparison of static, dynamic, and hybrid analysis for malware detection. Journal of Computer Virology and Hacking Techniques, 13 (1) 1-12 (2017).

[44] S. Morgan, (2017). Cyber security Business Report. Retrieved from CSO: https://www.csoonline.com/article/3237674/ransomware/ransomware-damage-costs-predicted-to-hit-115b-b y-2019.html.

[45] R. Collier, (2017). NHS ransomware attack spreads worldwide. CMAJ, 189 (22), 786-787. https://doi.org/10.1503/cmaj.1095434.

[46] H. Trisnasenjaya, I. Riadi Forensic Analysis of Android-based Whats App Messenger Against Fraud Crime Using The National Institute of Standard and Technology Framework. International Journal of Cyber Security and Digital Forensics (IJCSDF) 8 (1): 89-97 (2019).

[47] H. Parag Rughani. Artificial Intelligence Based Digital Forensics Framework. International Journal of Advanced Research in Computer Science. 8 (8): 10-14 (2017).

[48] 2016: Current State of Cybercrime, RSA Whitepaper, 2016.

[49] World Internet Users and 2017 Population Stats, accessed from http://http://www.internetworldstats.com/stats.htm.

[50] R. Mark. Computer forensics: Basics. Lecture note Purdue University (2004).

[51] Ibrahim Goni & Ahmed L. (2015) Propose Neuro-FuzzyGenetic Intrusion Detection System International Journal of Computer Applications Vol. 115 No. 8 available online at http://www.ijcaonline.com/archives/volume115/number8/20169-2320.