Attacking NGO Research with Pseudoscience

Nestori Syynimaa[0000-0002-6848-094X]

Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland
nestori.syynimaa@gmail.com

Abstract. Online attacks and harassment have increasingly targeted NGOs during the past few years. In this paper, we studied a recent disinformation campaign targeting Amnesty International and Citizen Lab Pegasus spyware research. We analyzed seven non-peer-reviewed reports published on ResearchGate claiming to scientifically prove Pegasus research unreliable. We identified five common claims in the reports and established all based on logical fallacies. Moreover, we identified multiple pseudoscience techniques used in the campaign, such as conspiracy theories, pretencing scientific support, and fierce attacks on legitimate scientists.

Keywords: disinformation, harassment, pseudoscience, non-government organization

1 Introduction

Online attacks and harassment targeting non-governmental human rights organizations (NGOs) have increased during the past few years[1]. Since 2016, Amnesty International (AI) and Citizen Lab (CL) have published multiple reports of governments using Pegasus software to spy on activists and journalists. Pegasus is a spyware software created and marketed by an Israelian NSO Group[2]. After publishing a report called "CatalanGate" [3] in 2022, CL and AI have been targeted with a disinformation campaign. Usually, digital authoritarians use their resources for mass manipulation and intimidation to silence the targets[1]. This campaign uses pseudoscience to question the work of CL and AI, and their attribution to both Pegasus and governments using it.

The campaign consists of multiple reports published on ResearchGate. According to the authors, "[t]he tools used to positively identify an NSO Pegasus spyware infection have been scientifically proven to be an unreliable and a dangerous source that can be easily manipulated" [4]. In this study, we will review these reports and analyze the claims made by the authors.

2 Research Methodology

We analyzed the reports used during the campaign listed in **Table 1** using qualitative content analysis [5] to identify common claims made by the authors. These claims were

analyzed to determine possible scientific and logical fallacies. Further, we examined common pseudoscientific techniques [6] used in the reports.

Table 1. Analyzed disinformation campaign reports

#	Year	Title	Author(s)
1	2023	PSUEDOSCIENCE. The Spyware Case of Omar Radi[7]	Scott, Jonathan
2	2023	CATALANGATE VECTORS: An Analysis of WhatsApp's Impact on Citizen Privacy and Amnesty International's MVT-Tool[8]	Scott, Jonathan Martin, Gregorio
3	2023	DISPROVING THE CITIZEN LAB & THEIR EXPERT[9]	Scott, Jonathan
4	2023	Exonerating Morocco EXONERATING MOROCCO DISPROVING THE SPYWARE[10]	Scott, Jonathan
5	2022	Exonerating Rwanda The Spyware Case of Carine Kanimba[11]	Scott, Jonathan
6	2022	Review of Catalangate Amnesty International Validation[12]	Scott, Jonathan
7	2022	UNCOVERING THE CITIZEN LAB -AN ANALYTICAL AND TECHNICAL REVIEW DISPROVING CATALANGATE [13]	Scott, Jonathan

3 Results

In this section, we will cover attacks against the work of CL and AI regarding Pegasus research. The disinformation campaign started after CL published a report called "CatalanGate", where Pegasus was used against Catalonian members of parliament [3]. After that, also previous reports covering Pegasus usage for Rwandan [14] and Moroccon [15] journalists and human rights activists were targeted.

In this section, we will show how the recent campaign is focused on certain repeating tenets: (1) Indications of Compromise (IoC), AI and CL research methodology, and (3) *ad hominem* attacks. Due to a lack of space, *ad hominem* attacks were left out of the scope of this report.

The reports listed in **Table 1** were all published on ResearchGate[16], a community platform where researchers can share their research for free. However, none of the papers was *published in peer-reviewed forums*, which is typical for pseudoscience practitioners [6]. Moreover, using ResearchGate as a publishing platform are example of *pretencing of scientific support* and *appealing directly to the public* [6].

3.1 Indication of Compromise

AI Security Lab has created a Mobile Verification Toolkit (MVT), which is a set of utilities for gathering forensic traces from potentially compromised Android and iOS devices. It has been developed as a part of the Pegasus project and is meant for "technologists and investigators", not "for end-user self-assessment" [17].

The tool has two options for acquiring the data to be analyzed; *Filesystem Dump* and *iTunes backup*. The former requires jailbreaking the iPhone, which can cause

unintended malfuctioning of the device and will likely void the warranty of the device. The latter option will generate an iTunes backup, which is a non-destructive method and fully supported by iPhone manufacturer Apple. The backups do not have all the data stored in the device. Still, they are "useful in cases where other acquisition types are not feasible" [18, p. 77] and in many cases, "sufficient to at least detect some suspicious artifacts" [19].

Indication of Compromise (IoC) can be defined as a "technical artifact or observable that suggests an attack is imminent or is currently underway, or that a compromise may have already occurred." [20]. As such, IoC can be any evidence indicating threat actor activities, such as URLs, log entries, files, etc. *False positive* refers to an IoC that is detected but not malicious, and *false negative* to an IoC that is malicious but not detected. The quality of detecting IoCs is measured by *precision* (purity of retrieval) and *recall* (completeness of retrieval) [21]. In other words, good precision means very few false positives and good recall, very few false negatives.

IoC should be based on Tactics, Techniques, and Procedures (TTP) threat actors use [22]. Developing IoCs is similar to a scientific process: IoCs are derived based on what is learned by analyzing TTPs, tested, and revised as needed. Generally, IoCs are not the best available data, as they "fails to identify novel or changing threats that don't match known indicators, and only provides detection capabilities after the fact." [22, p. 5]. However, for tools like MVT that analyze historical data, the only available evidence are IoCs.

Good recall is more important than good precision for IoCs related to spying. It ensures that as many as possible infected devices are found. False positives are not an issue, as individual IoCs are validated by forensic analysts [23]. As such, IoCs are input for forensics analysis, not the end of it.

Reported IoCs can't be trusted. IoCs can be derived using static, dynamic, or hybrid analysis [24]. Technical details on analyzing Pegasus are published in multiple reports [25-27]. Regardless, claims like "Amnesty and Citizen Lab have failed to provide TTPs for the alleged hacked devices. This lack of information makes it easy to forge IOCs and results in disputes within their information science." [8, p. 11] and "[research methodology] relied upon mere conjecture, failing to provide the essential supporting evidence required in this line of research" [7] are made. These claims can be summarized as follows:

> Claim 1. AM and CL have not shared details on how they derived IoCs. Therefore, IoCs are fake.

This claim is an example of *argumentum ad ignorantiam*, which means that lack of evidence does not prove the result wrong [28]. Also, the first claim suggests that AI or CL may have forged IoCs, which is an example of a *conspiracy theory* [6].

Another campaign target has been certain classes of IoCs, namely domain names and IP addresses. For instance, CL attributed domains *nnews[.]co* and *123tramites[.]com* to Pegasus, as they "were complete matches for our fingerprint" [3, p. 20].

To validate the CL forensic methods, AI independently verified three CatalanGate cases "using their own forensic methodology" [3, p. 23]. This statement has been

targeted by multiple campaign reports [e.g., 8, 12, 13]. For instance, "[t]he verification by Amnesty of 123tramites[.]com would not have been possible because it had already been expired for 6 months, and there was no Pegasus exploit server to verify as Citizen Lab claims in their report. The same issue of credible verification by Amnesty International arises for all other alleged domain name indicators of compromise." [13, p. 26]. Here the attackers are *fabricating fake controversy* [6] by claiming AI could not have verified the mentioned IoCs. AI never contended that they confirmed the specific IoCs, but Pegasus *infections*. Moreover, to prevent researchers from finding the location of Pegasus exploit servers, "[e]ach subdomain was generated and only active for a short period of time" [27, p. 31], which would have made confirming the existence of these servers impossible in the first place. As such, the claim is another example of *argumentum ad ignorantiam*. These claims can be summarized as follows:

Claim 2. IoC domains were not active during AI verification. Therefore, IoCs are fake.

MVT Tool finds fake positives. As mentioned, MVT tool is built for investigators to collect data to detect possible Pegasus infections. The campaign has also targeted this tool [e.g.,8, 12, 13]. For instance, "[a]fter reading through the code in the MVT-Tool it was easy to determine that the tool used to detect if a mobile device is infected with spyware is nothing more than a search for keywords. The keywords used to search for the infection are derived from the indicators of compromise published by Citizen Lab and Amnesty International." [13, p. 30]. Indeed, the MVT tool is used to recognize IoCs, such as domain and file names. To test the reliability of the MVT tool, attackers conducted four experiments.

In the first experiment, nine people were asked to create WhatsApp messages containing Pegasus IoC domains and send them to themselves [29]. One out of nine participants could not make iTunes backup due to memory issues. The MVT tool found the spoofed IoCs for the rest of the participants as expected. However, this was reported as 88,9 per cent of participants yielding a false positive infection. This is another example of *pretencing to have support in science*.

In the second experiment, process IoCs were added to *Manifest.db*[12], used by the MVT tool. Again, the MVT tool found the spoofed IoCs as expected. This was reported as "I knew that I had full reign to spoof any indicator of compromise I wanted" [12, p. 25].

In the third experiment, an iPhone without internet connection was used to browse to Pegasus IoC domains [8]. Again, MVT tool found the spoofed IoCs. This was reported as “MVT-Tool finds all 7 addresses to be positive for Pegasus without ever having an internet connection” [8, p. 12].

In the fourth experiment, an iPhone application was created using *DiagnosticD* as a bundle identifier [30], one of the Pegasus IoCs. Again, the MVT tool found the spoofed IoCs as expected. These experiments are an example of the logical fallacy of *affirming the consequent* [31], and the claim can be summarized as follows:

Claim 3. MVT-tool found spoofed false positive IoCs. Therefore, all found IoCs are fake.

Other IoC-related attacks targeted real false positives—for instance, the false positive IoC *com.apple.CrashReporter.plist* was removed by AI. The attacker reported this as "it creates a false positive regressively nullified every case that included *com.apple.CrashReporter.plist* before Jan 12th, 2022" [12, p. 26]. However, as the AI report shows [27, appx. D], in most cases, this IoC was one of many. This claim is another example of *affirming the consequent* and can be summarized as follows:

Claim 4. MVT-tool found false positive IoC from the device. Therefore, all found IoCs are fake.

3.2 Research Methodology

During the campaign, attackers compared[7, 12] the methodology used in AI and CL reports to European Union Agency for Network and Information Security (ENISA) guide on electronic evidence [32]. The guide provides principles for electronic evidence gathering, including an audit trail, which has been targeted during the campaign. For instance, "[h]ow did The Citizen Lab conduct their investigations without performing a logical or physical data analysis on the device?" [11, p. 25]. Moreover, "[w]ithout ever having the mobile device and working with backup that were not taken by them, and can easily be tampered with." [12, p. 9]. These claims are yet another example of *affirming the consequent* and can be summarized as follows:

Claim 5. iTunes backups can be tampered with. Therefore, all iTunes backups are fake.

4 Summary

We identified multiple techniques used in the campaign that are typical of pseudoscience. For instance, the inability to publish in peer-reviewed media, fabrication of fake controversies, pretencing of scientific support, conspiracy theories, appeals directly to the public, fierce attacks on legitimate scientists, and strong political connections were identified throughout the campaign.

The alleged scientific claims used during the disinformation campaign lacked the logic and proof required from scientific claims. As such, the campaign reports do not provide any claimed evidence to undermine AI and CL Pegasus research. The summary of the campaign claims and logical fallacies are listed in **Table 2**.

Table 2. Campaign claims

#	Claim	Logical fallacy
1	AM and CL have not shared details on how they derived IoCs. Therefore, IoCs are fake	argumentum ad ignorantiam
2	IoC domains were not active during AI verification. Therefore, IoCs are fake	argumentum ad ignorantiam

#	Claim	Logical fallacy
3	MVT-tool found spoofed false positive IoCs. Therefore, all found IoCs are fake	affirming the consequent
4	MVT-tool found false positive IoC from the device. Therefore, all found IoCs are fake	affirming the consequent
5	iTunes backups can be tampered with. Therefore, all iTunes backups are fake	affirming the consequent

References

1. Heller, B., Enlisting Useful Idiots: The Ties Between Online Harassment and Disinformation. Colorado Technology Law Journal, 2021. **19**(1): p. 19-42.
2. NSO Group. NSO Group website. 2023 [May 6th 2023]; Available from: https://www.nsogroup.com/.
3. Scott-Railton, J., et al., CatalanGate: Extensive Mercenary Spyware Operation against Catalans Using Pegasus and Candiru. 2022, University of Toronto: Toronto.
4. Scott, J. NSO THROUGH THE VEIL. 2022 [May14th 2023]; Available from: https://jonathandata1.medium.com/nso-through-the-veil-ce490fd862f4.
5. Cho, J.Y. and E.-H. Lee, Reducing confusion about grounded theory and qualitative content analysis: Similarities and differences. Qualitative report, 2014. **19**(32).
6. Hansson, S.O., Science denial as a form of pseudoscience. Studies in History and Philosophy of Science Part A, 2017. **63**: p. 39-47.
7. Scott, J. PSUEDOSCIENCE. The Spyware Case of Omar Radi. 2023 [30]. Available from: https://www.researchgate.net/publication/371313767_Pseudoscience_-_The_Spyware_Case_of_Omar_Radi.
8. Scott, J. and G. Martin. CATALANGATE VECTORS: An Analysis of WhatsApp's Impact on Citizen Privacy and Amnesty International's MVT- Tool. 2022 [21]. Available from: https://www.researchgate.net/profile/Jonathan-Scott-26/publication/367078083.
9. Scott, J. DISPROVING THE CITIZEN LAB & THEIR EXPERT. 2023 [9]. Available from: https://www.researchgate.net/profile/Jonathan-Scott-26/publication/369022555.
10. Scott, J. Exonerating Morocco EXONERATING MOROCCO DISPROVING THE SPYWARE. 2023 [27]. Available from: https://www.researchgate.net/profile/Jonathan-Scott-26/publication/368607677.
11. Scott, J. Exonerating Rwanda The Spyware Case of Carine Kanimba. 2022 [40]. Available from: https://www.researchgate.net/profile/Jonathan-Scott-26/publication/366298195.
12. Scott, J. Review of Catalangate Amnesty International Validation. 2022 [46]. Available from: https://www.researchgate.net/profile/Jonathan-Scott-26/publication/365743925.
13. Scott, J. UNCOVERING THE CITIZEN LAB -AN ANALYTICAL AND TECHNICAL REVIEW DISPROVING CATALANGATE. 2022 [59]. Available from: https://www.researchgate.net/publication/361738419.
14. Amnesty International. Moroccan Journalist Targeted With Network Injection Attacks Using NSO Group's Tools. 2020 [Available from: https://www.amnesty.org/en/latest/research/2020/06/moroccan-journalist-targeted-with-network-injection-attacks-using-nso-groups-tools/.
15. Amnesty international. Morocco: Human Rights Defenders Targeted with NSO Group's Spyware. 2019 [May 6th 2023]; Available from: https://www.amnesty.org/en/latest/research/2019/10/Morocco-Human-Rights-Defenders-Targeted-with-NSO-Groups-Spyware/.

16. ResearchGate. Research Gate. 2023 [Jun 16th 2023]; Available from: https://www.researchgate.net/.
17. Amnesty International Security Lab. github: Mobile Verification Toolkit. 2021 [Available from: https://github.com/mvt-project/mvt.
18. Tamma, R., et al., Practical mobile forensics: Forensically investigate and analyze iOS, Android, and Windows 10 devices. 2020: Packt Publishing Ltd.
19. MVT Project Developers. iOS Forensic Methodology. 2022 [May 2nd 2023]; Available from: https://docs.mvt.re/en/latest/ios/methodology/.
20. NIST. Computer Security Resource Center. Glossary: Indicator. 2023 [Available from: https://csrc.nist.gov/glossary/term/indicator.
21. Buckland, M. and F. Gey, The relationship between recall and precision. Journal of the American society for information science, 1994. **45**(1): p. 12-19.
22. Daszczyszak, R., et al., TTP-Based Hunting, in MITRE Technical Report. 2019.
23. Machaka, V. and T. Balan, Investigating Proactive Digital Forensics Leveraging Adversary Emulation. Applied Sciences, 2022. **12**(18): p. 9077.
24. Akram, B. and D. Ogi. The Making of Indicator of Compromise using Malware Reverse Engineering Techniques. in 2020 International Conference on ICT for Smart Society (ICISS). 2020.
25. Bazaliy, M., et al., Technical Analysis of the Pegasus Exploits on iOS. 2016.
26. Bazaliy, M., et al., Technical Analysis of Pegasus Spyware. An Investigation Into Highly Sophisticated Espionage Software. 2016. p. 35.
27. Amnesty International, Forensic Methodology Report: How to Catch NSO Group's Pegasus. 2021, Amnesty International: London, UK.
28. Pigliucci, M. and M. Boudry, Prove it! The Burden of Proof Game in Science vs. Pseudoscience Disputes. Philosophia, 2014. **42**(2): p. 487-502.
29. Scott, J. github: Pegasus-CatalanGate-False-Positives. 2022 [May 2nd 2023]; Available from: https://github.com/jonathandata1/Pegasus-CatalanGate-False-Positives.
30. Scott, J. github: Legitimate Apple Apps can be seen as malicious - False Positive Results #320. 2022 [Available from: https://github.com/mvt-project/mvt/issues/320.
31. Jevons, W.S., *Elementary lessons in logic: Deductive and inductive: With copious questions and examples, and a vocabulary of logical terms*. 1872: Macmillan.
32. ENISA, *Electronic evidence - a basic guide for First Responders. Good practice material for CERT first responders*. 2015. p. 26.