




A Review of Eight Reasons Why Cybersecurity on Novel Generations of Brain-Computer Interfaces Must Be Prioritized

Sergio López Bernal ¹, Alberto Huertas Celdrán², Gregorio Martínez Pérez¹

¹*Department of Information and Communications Engineering, University of Murcia, 30100 Murcia, Spain*
 {slopez, gregorio}@um.es

²*Communication Systems Group CSG, Department of Informatics IfI, at the University of Zurich UZH, CH 8050 Zürich, Switzerland*
 huertas@ifi.uzh.ch

Abstract—Brain-computer interfaces enable bidirectional communication between the brain and external devices, used in medicine for diagnosing and treating neurodegenerative diseases. Despite their advantages, they present vulnerabilities that attackers could exploit to cause brain damage. In this context, previous work defined three neural cyberattacks altering spontaneous neuronal activity. However, more effort is still needed to detect and characterize new neural cyberattacks with new behaviors. Based on that, this publication presents a taxonomy of eight neural cyberattacks, where five of them are novel. For each of them, this work offers a formal definition and the conceptualization of their behavior. Finally, it compares their impact on the short and long term, indicating that Neuronal Nonce was the most damaging attack in the short term, with an approximate 12% reduction of neural activity compared to spontaneous behavior. Finally, Neuronal Scanning was the most effective in the long term, offering a reduction of around 9%.

Index Terms—BCI, cybersecurity, neural cyberattacks, brain

Tipo de contribución: *Investigación ya publicada*

I. INTRODUCTION

Brain-Computer Interfaces (BCIs) represent bidirectional systems interacting with the brain, enabling neural data acquisition and neuronal stimulation. These interfaces vary in invasiveness, where invasive interfaces are widely utilized in medical scenarios. For instance, invasive BCIs focusing on neural recording have facilitated the control of prosthetic limbs in patients with impairments, while those geared towards neuromodulation can treat neurodegenerative conditions like Parkinson’s disease.

Despite the benefits of novel invasive BCIs, the literature has documented vulnerabilities susceptible to exploitation. Notably, two cyberattack techniques targeting neural stimulation, namely Neural Flooding and Neural Scanning, along with a neural inhibition-focused cyberattack, have been identified. Known as neural cyberattacks, these threats can disrupt the spontaneous activity of brain neural networks by stimulating or inhibiting neurons. However, the literature lacks formal definitions of these cyberattacks’ behaviors, as well as how they might emulate effects seen in specific neurodegenerative diseases. Moreover, the study of the impact generated by these cyberattacks on spontaneous neural activity remains an open challenge.

This work summarizes the research published in [1], proposing a taxonomy of eight neural cyberattacks affect-

ing spontaneous neural activity, inspired by cyberattacks in the computer science realm. Following the presentation of their formal definitions, these cyberattacks were applied to a simulated biological neural network modeling a section of a mouse’s visual cortex. This paper also evaluates the short and long-term effects of the proposed neural cyberattacks, facilitating a comparative analysis of their impacts.

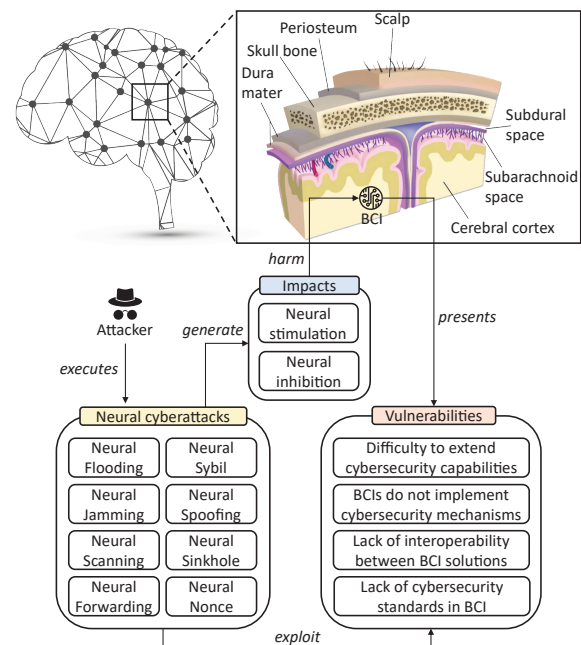


Figure 1. Attacker executing the proposed neuronal cyberattacks that exploit vulnerabilities of invasive neuromodulation BCIs and impact the BCI.

II. EIGHT NEURAL CYBERATTACKS AFFECTING BRAIN BEHAVIOR

This section highlights the main aspects of each cyberattack proposed. First, Neuronal Flooding (FLO) aims to stimulate a set of neurons in a specific instant. As an example, Figure 2 presents the overall behavior of the FLO attack implemented. In contrast, Neuronal Jamming (JAM) inhibits the activity of a set of neurons during a temporal window, making them unable to generate or transmit impulses to near neurons.

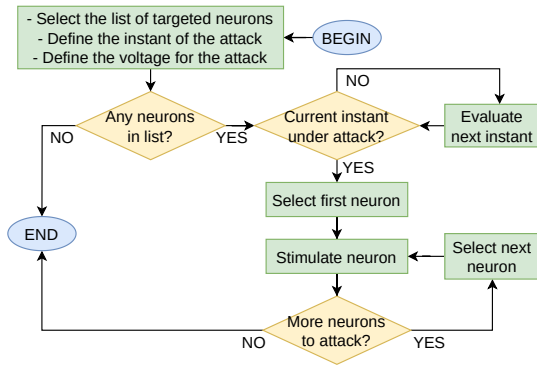


Figure 2. Implemented behavior of Neuronal Flooding.

Neuronal Scanning (SCA) cyberattacks sequentially stimulate all neurons within a neuronal population, targeting one neuron per instant. Neuronal Selective Forwarding (FOR) alters the propagation of a set of neurons during a temporal window, inhibiting specific neurons for each instant of the window.

Furthermore, Neuronal Spoofing (SPO) consists in reproducing the behavior of a set of neurons during a given period, consisting first on recording neural activity to then recreate the obtained pattern within the same or different neurons during a different window. Regarding Neuronal Sybil (SYB), an attacker could change the behavior of one or more neurons to do the opposite as their natural behavior (inhibit when the neuron would fire, or stimulate when it was not firing). Neuronal Sinkhole (SIN) focuses on stimulating neurons from superficial layers connected to neurons located in deeper layers, being the later the main focus of the attack. Finally, Neuronal Nonce (NON) aims to attack a random set of neurons in a specific instant. The action enforced could change according to the interests of the attacker, generating neural stimulation, neural inhibition, or a combination of both.

III. IMPACT OF NEURAL CYBERATTACKS

During this research the literature lacked realistic neuronal topologies. Consequently, a simulated biological network was utilized, generated artificially by training a Convolutional Neural Network (CNN) to address the particular problem of a mouse that has to exit a maze of size seven by seven positions, with certain obstacles. The weights obtained from this CNN were transformed into biological synaptic weights, serving to model the voltage increase triggered during neuronal activation. This methodology was chosen due to the similarities observed between CNNs and the cerebral visual cortex in the literature, both in terms of their incremental function and architectural characteristics. In this context, training the CNN concluded an optimal path of 27 positions to find the maze exit from the starting position, which was later used to provide a 27-second neuronal simulation where the mouse stayed one second per position of the optimal path.

Figure 3 illustrates the impact of each cyberattack relative to spontaneous behavior indicating the percentage reduction of spikes. This figure indicates a distinction between the first five positions and the last five positions along the optimal path within the maze to find the exit, highlighting which

cyberattacks represent more immediate harm and which are better suited for prolonged threats. The variability observed per cyberattack corresponds to differences among the five positions considered, whether initial or final. Furthermore, for FLO, JAM, and SYB, which employ random selection of target neurons, ten executions were conducted to introduce variability. In contrast, the results for NON only contain one execution given its inherent randomness.

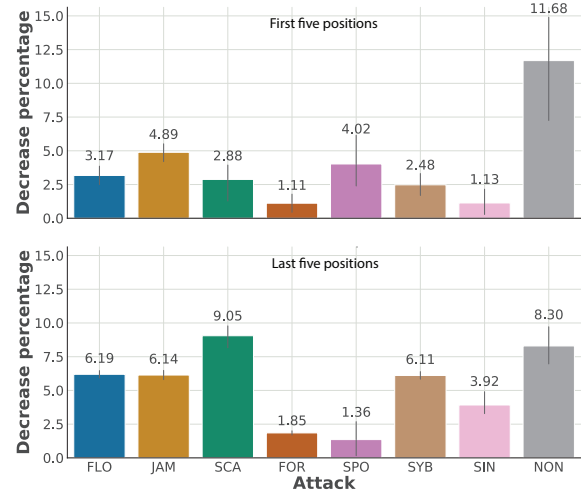


Figure 3. Mean percentage of spikes reduced per neural cyberattack compared with spontaneous behavior, studied over the first and last five positions of the maze in a biological simulation of 27 seconds.

According to the results, NON achieves a 12% reduction in spontaneous activity within the initial five positions due to its natural randomness, making it the most damaging cyberattack in the short term, followed by JAM with an almost 5% reduction. Oppositely, SCA comes as the most impacting attack for the long term, inducing a spike reduction of approximately 9%, followed by NON with an 8% reduction.

IV. CONCLUSION

This work presents a taxonomy of eight neural cyberattacks capable of disrupting spontaneous neural activity by inducing neuronal stimulation or inhibition. It defines two groups of cyberattacks, either performing the attack at a certain instant or during a temporal window, evaluated on a simplified neuronal topology representing a portion of a mouse's visual cortex. Future work could explore the use of more realistic neuronal topologies, the detection of neural cyberattacks, and their applicability over different cognitive functions.

V. ACKNOWLEDGEMENTS

This work has been supported by the strategic project CDL-TALENTUM from the Spanish National Institute of Cybersecurity (INCIBE) and by the Recovery, Transformation and Resilience Plan, Next Generation EU.

REFERENCES

- [1] S. López Bernal, A. Huertas Celdrán, and G. Martínez Pérez, "Eight reasons to prioritize brain-computer interface cybersecurity," *Communications of the ACM*, vol. 66, no. 4, p. 68–78, mar 2023.