Apuntes de Taxonomía de Ciberataques Neuronales

Índice

1. Resumen

El artículo presenta una taxonomía de ocho ciberataques neuronales dirigidos a las nuevas generaciones de interfaces cerebro-computadora (BCI), centrándose en su capacidad para alterar la actividad neuronal espontánea mediante la estimulación o inhibición maliciosa. Inspirados en ataques tradicionales de la informática (por ejemplo, inundación, bloqueo, suplantación), estos ciberataques explotan vulnerabilidades en BCIs tanto invasivas como no invasivas, como las utilizadas en terapias médicas (por ejemplo, Neuralink). Los autores simulan estos ataques en una red neuronal biológica modelada a partir de la corteza visual de un ratón, generada utilizando una red neuronal convolucional (CNN) debido a la falta de topologías neuronales realistas. Los hallazgos clave revelan que el **Neural Nonce** (estimulación/inhibición aleatoria) y el **Neural Jamming** (inhibición temporal) son los más impactantes a corto plazo, reduciendo los picos neuronales en 12% y 5%, respectivamente. Para efectos a largo plazo, el **Neural Scanning** (estimulación secuencial) y el **Neural Nonce** muestran la mayor reducción de picos (9% y 8%). El estudio subraya los riesgos críticos de ciberseguridad en las neurotecnologías emergentes, enfatizando la necesidad de defensas robustas para prevenir ataques que podrían imitar o exacerbar condiciones neurodegenerativas. Estos resultados resaltan la urgencia de esfuerzos interdisciplinarios para asegurar las BCIs a medida que avanzan hacia una adopción clínica y de consumo más amplia.

2. Ciberataques Neuronales

Se van a explicar los ocho ciberataques neuronales explorados en el artículo.

2.1. Neuronal Flooding (FLO)

- **Descripción:** En cibersguridad, un ataque FLO se basa en colapsar una red enviando una gran cantidad de paquetes. Traduciendo esto a una perspectiva neurológica, el «Neuronal Flooding» trata de estimular simultáneamente un gran número de neuronas.
- **Complejidad:** Baja. No requiere conocer el estado de las neuronas de la red objetivo.

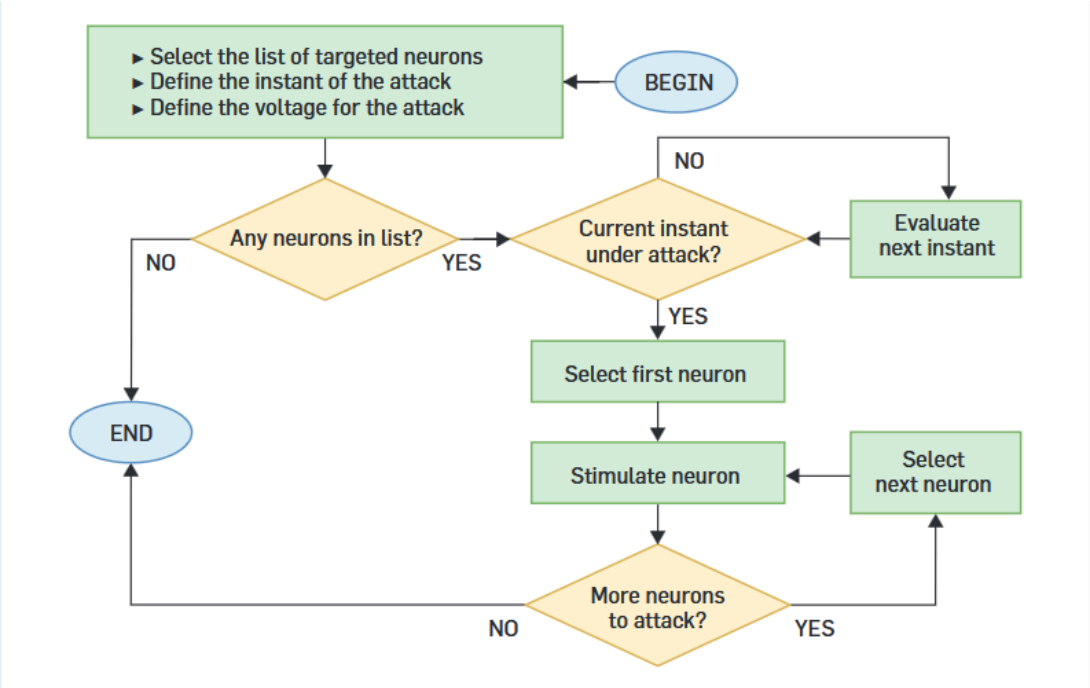

Figura 1: Esquema de funcionaiento del ataque Neuronal Flooding (FLO).

2.2. Nuronal Jamming (JAM)

- **Descripción:** En ciberseguridad, el ataque JAM introduce interferencias maliciosas para bloquear la comunicación. En el contexto neuronal, el «Neuronal Jamming» consiste en inhibir durante una ventana temporal la excitación de un conjunto de neuronas.
- **Complejidad:** Baja. No requiere conocer el estado de las neuronas de la red objetivo.

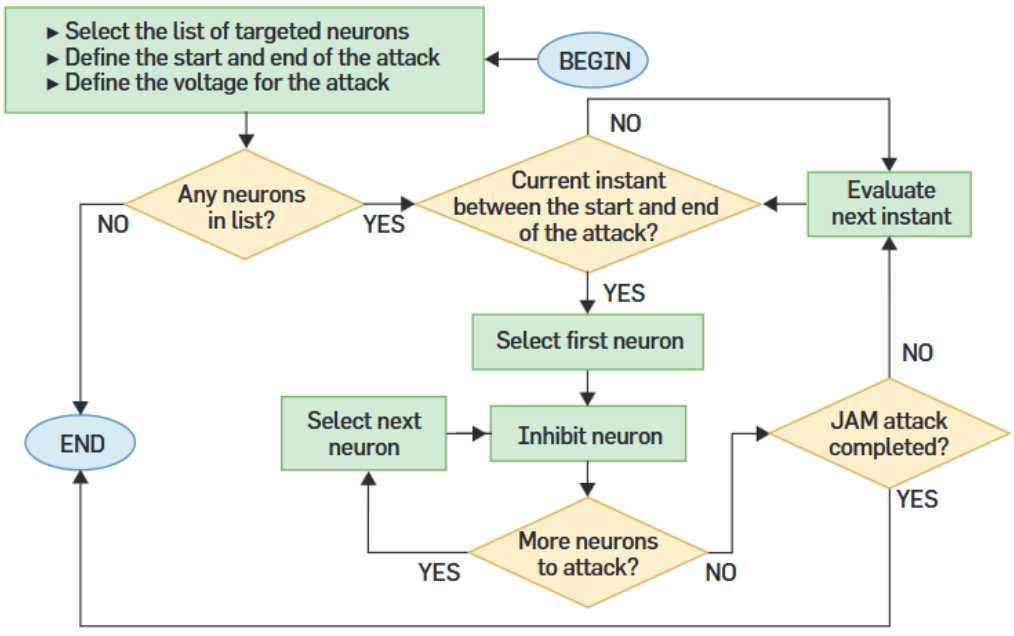

Figura 2: Esquema de funcionaiento del ataque Neuronal Jamming (JAM).

2.3. Neuronal Scanning (SCA)

- **Descripción:** En ciberseguridad, el ataque SCA comprueba la conexión de distintos dispositivos para encontrar vulnerabilidades. Para ello, escanea uno a uno los diferentes dispositivos. En el ámbito neuronal, el «Neuronal Scanning» consiste en estimular las neuronas de un conjunto una a una, de forma que en cada instante de tiempo solo se excita una sola neurona.
- **Complejidad:** Baja. No requiere conocer el estado de las neuronas de la red objetivo.

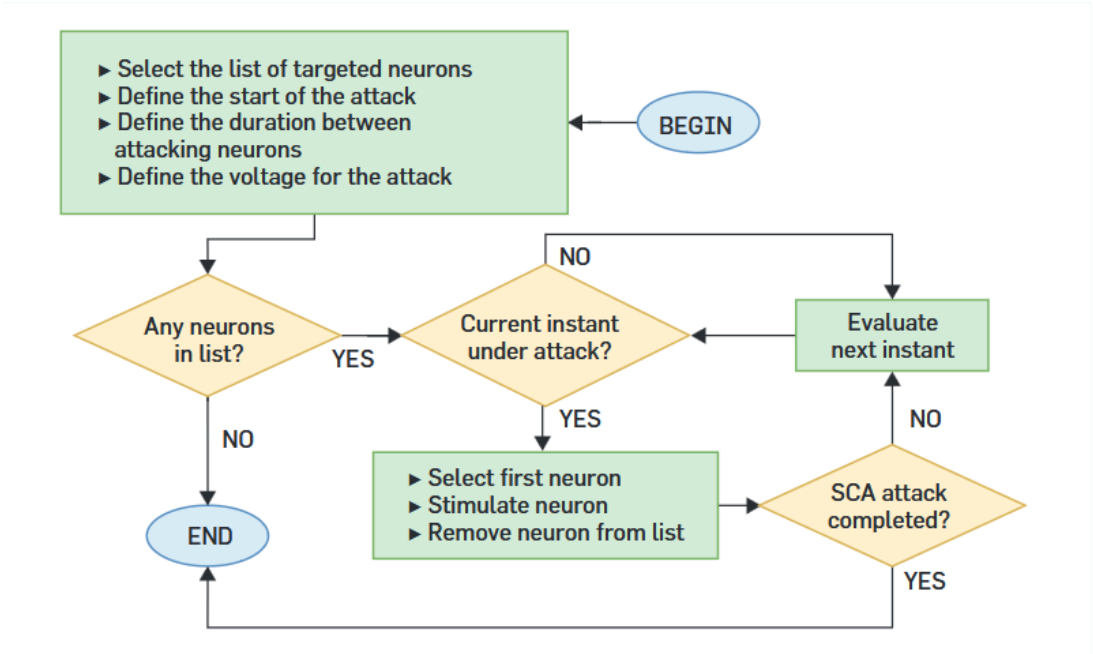

Figura 3: Esquema de funcionaiento del ataque Neuronal Scanning (SCA).

2.4. Neuronal Selective Forwarding (FOR)

- **Descripción:** En ciberseguridad, el ataque FOR pierde selectivamente paquetes de la red en vez de enviarlos. En el contexto neuronal, el «Neuronal Selective Forwarding» consiste en cambiar el comportamiento de un conjunto de neuronas en una ventana temporal, enhibiendo la excitación en cada instante de la ventana.
- **Complejidad:** Alta. Requiere conocer el comportamiento de las neuronas afectadas por el ataque, monitoreándolas en tiempo real.

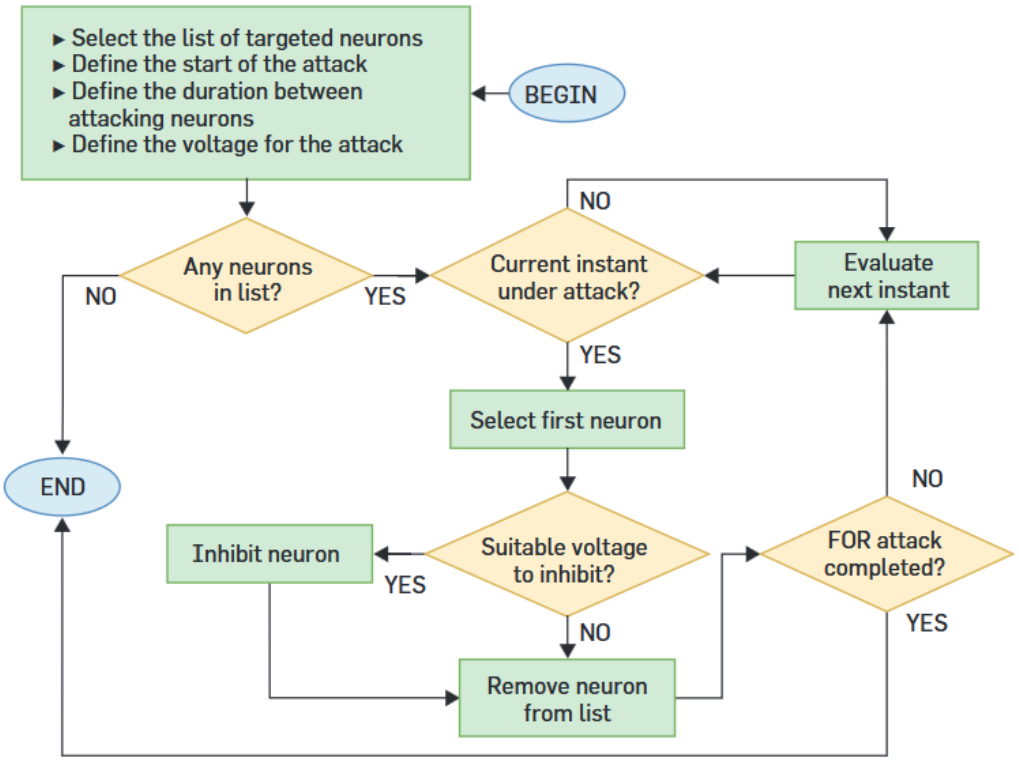

Figura 4: Esquema de funcionaiento del ataque Neuronal Selective Forwarding (FOR).

2.5. Neuronal Spoofing (SPO)

- **Descripción:** En ciberseguridad, el ataque SPO ocurre cuando una entidad maliciosa engaña a un sujeto para robarle información o para lanzar ataques contra otros «hosts». En el contexto neuronal, el «Neuronal Spoofing» consiste en replicar el comportamiento de un conjunto de neuronas durante un tiempo determinado.
- **Complejidad:** Muy alta. Requiere tanto grabar, como estimular e inhibir.

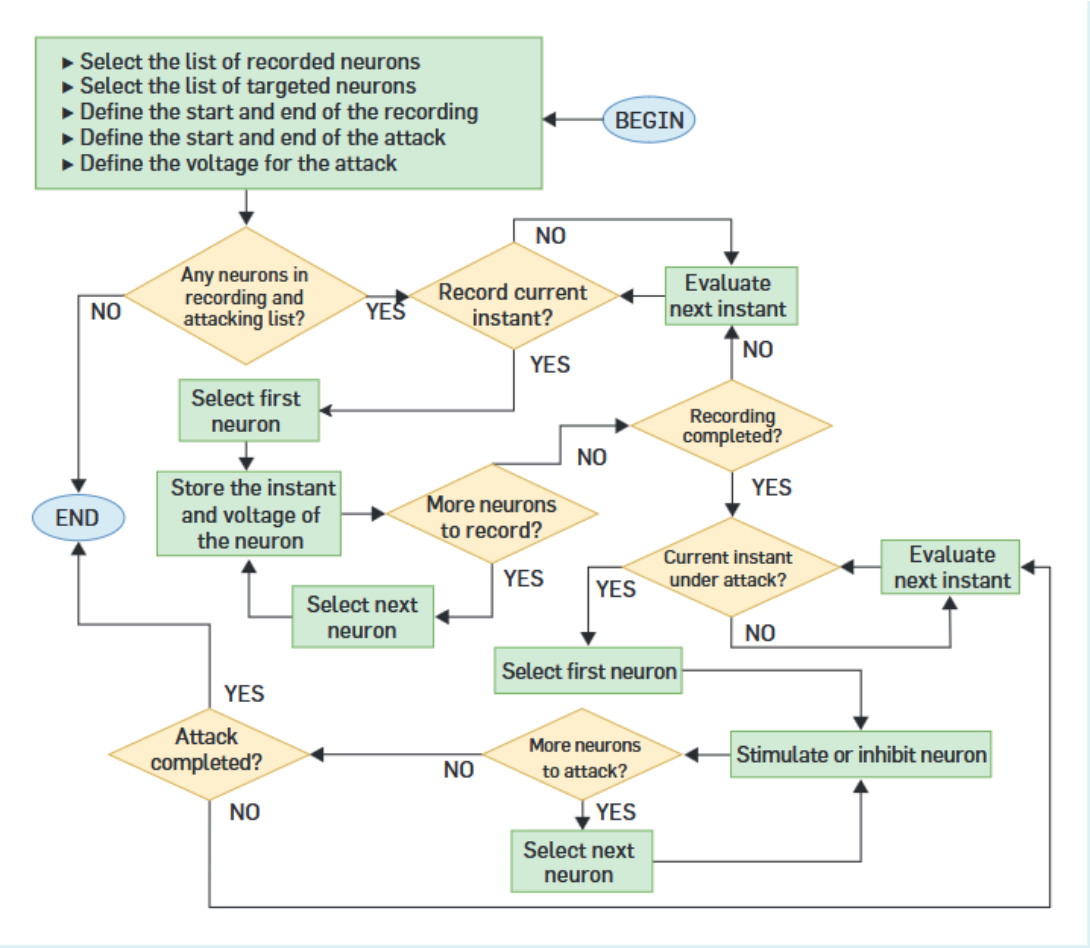

Figura 5: Esquema de funcionaiento del ataque Neuronal Spoofing (SPO).

2.6. Neuronal Sybil (SYB)

- **Descripción:** EN ciberseguridad, el ataque SYB consiste en crear múltiples identidades falsas para engañar a un sistema. En el contexto neuronal, el «Neuronal Sybil» consiste en modificar el comportamiento de una o más neuronas para que hagan justo lo contrario a lo que deberían.
- **Complejidad:** Muy alta. Requiere tanto grabar (en tiempo real), como estimular e inhibir.

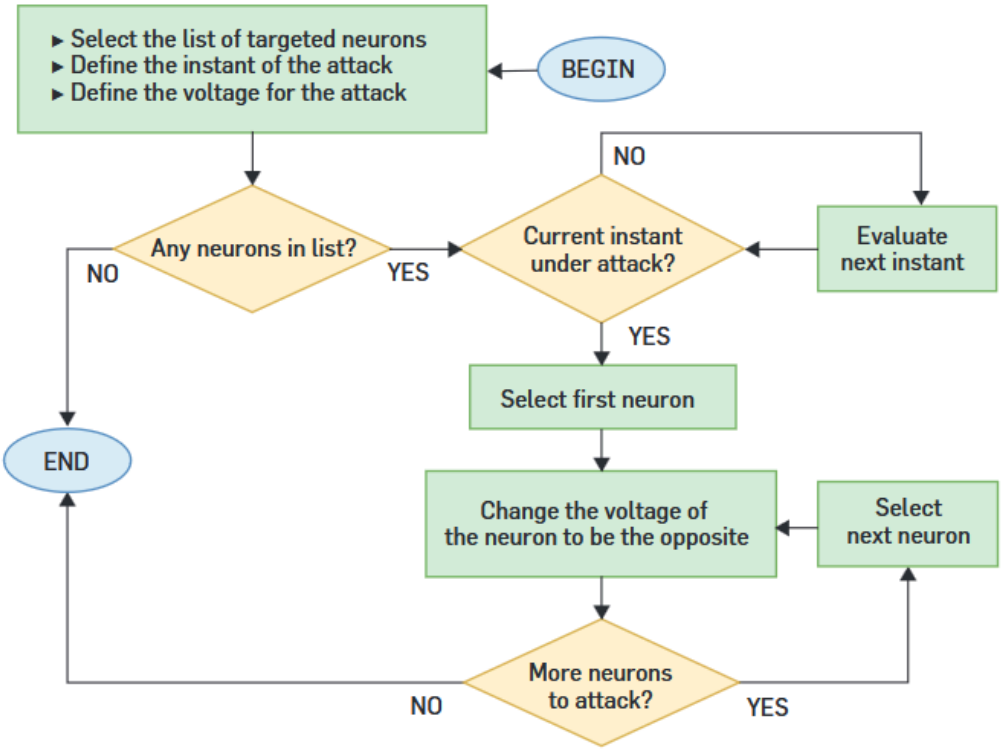

Figura 6: Esquema de funcionaiento del ataque Neuronal Sybil (SYB).

2.7. Neuronal Sinkhole (SIN)

- **Descripción:** En cibersguridad, el ataque SIN consiste en redirigir el tráfico de una red a un destino malicioso, de forma que el destino pueda hacer que lo quiera con la información. En el contexto neuronal, el «Neuronal Sinkhole» consiste en excitar un conjunto de neuronal superficiles que están conectadas a neuronas más profundas, redirigiendo así la actividad neuronal a un punto específico.
- **Complejidad:** Muy alta. Requiere conocer la topología de la red neuronal.

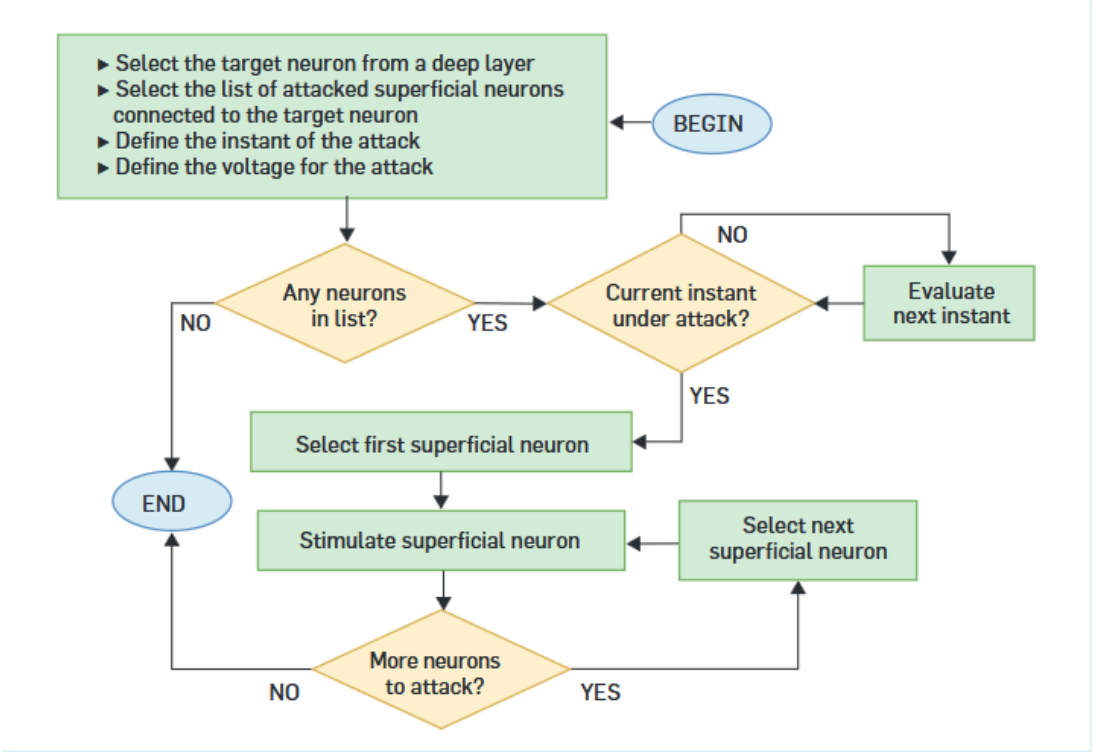

Figura 7: Esquema de funcionaiento del ataque Neuronal Sinkhole (SIN).

2.8. Neuronal Nonce (NON)

- **Descripción:** En ciberseguridad, «Nonce numbers» se refiere a valores aleatorios utilizados en criptografía. En el contexto neuronal, el «Neuronal Nonce» consiste en atacar a un conjunto aleatorio de neuronas en un instante concreto. El ataque a las neuronas varía aleatoriamente entre estimular, inhibir o no hacer nada.
- **Complejidad:** Baja. No requiere conocer el estado de las neuronas de la red objetivo.

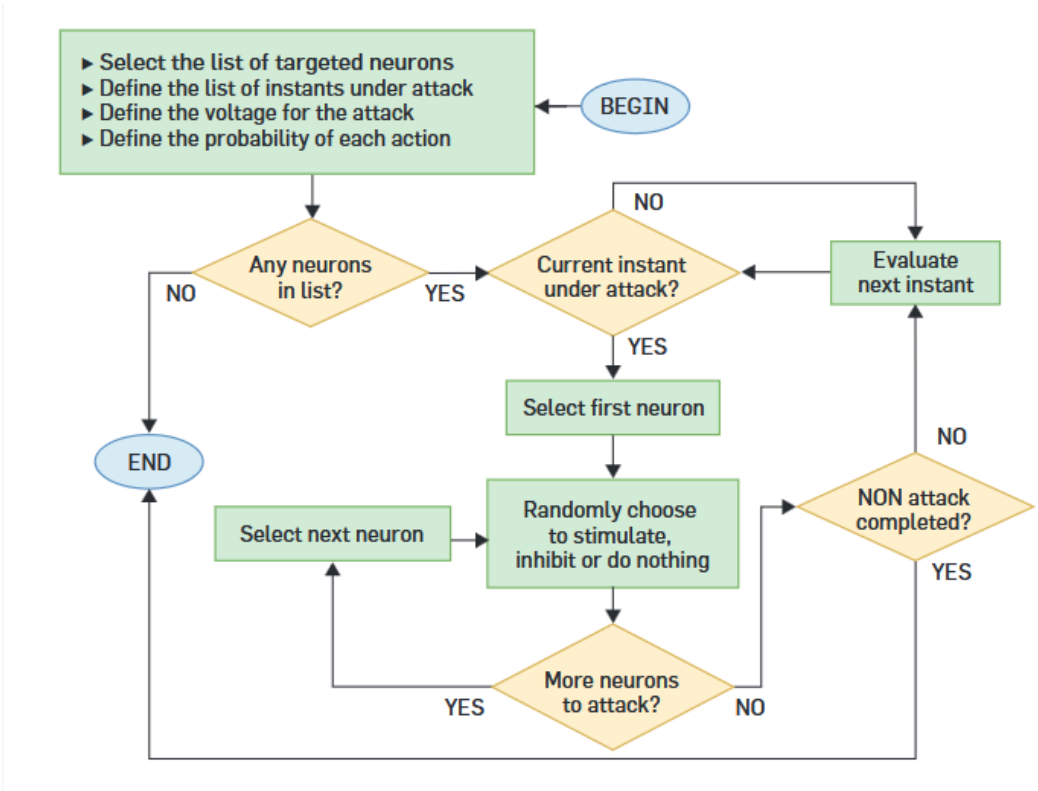

Figura 8: Esquema de funcionaiento del ataque Neuronal Nonce (NON).