# A deep learning technique to generate DNS amplification attacks

July 14 2022, by Ingrid Fadelli

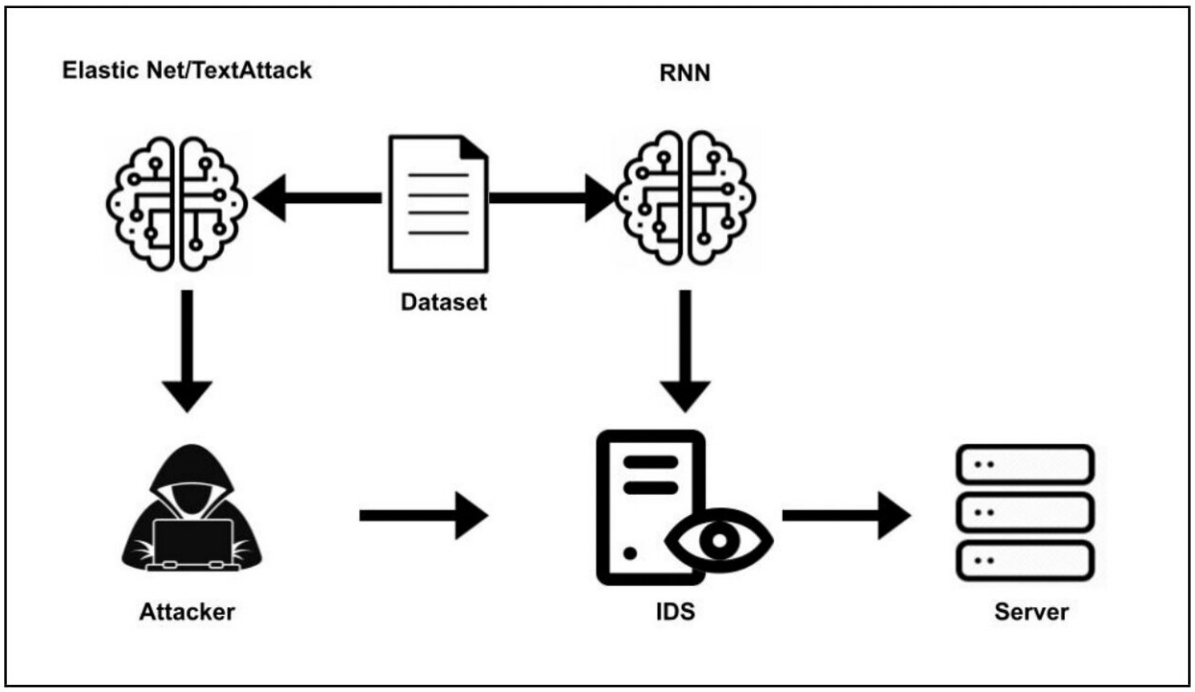


Diagram outlining the structure of the team's experiment. Credit: Mathews et al.

Deep learning techniques have recently proved to be highly promising for detecting cybersecurity attacks and determining their nature. Concurrently, many cybercriminals have been devising new attacks aimed at interfering with the functioning of various deep learning tools, including those for image classification and natural language processing.

Perhaps the most common among these attacks are adversarial attacks, which are designed to "fool" deep learning algorithms using data that has been modified, prompting them to classify it incorrectly. This can lead to the malfunctioning of many applications, biometric systems, and other technologies that operate through deep learning algorithms.

Several past studies have shown the effectiveness of different adversarial attacks in prompting deep neural networks (DNNs) to make unreliable and false predictions. These attacks include the Carlini & Wagner attack, the Deepfool attack, the fast gradient sign method (FGSM) and the Elastic-Net attack (ENA).

Researchers at the Citadel have recently developed a DNN that can detect a type of cyberattack known as distributed denial of service (DDoS) DNS amplification, and then used two different algorithms to generate adversarial examples that could trick their DNN. Their findings, published in a paper pre-published on arXiv, further confirm the unreliability of deep learning methods for DNS attack detection and their vulnerability to adversarial attacks.

DDoS DNS amplification attacks exploit vulnerabilities of domain name system (DNS) servers to amplify the queries made to them, ultimately flooding them with information and bringing the servers down. These attacks can cause significant disruption to online services, including those run by both small and big multinational companies.

Over the past few years, computer scientists have developed several deep learning techniques that can detect DDoS DNS amplification attacks. Nonetheless, the team at the Citadel showed that these techniques could be circumvented using adversarial networks.

"Much of the current work in the field of adversarial learning has been conducted in image processing and natural language processing with a

wide variety of algorithms," Jared Mathews and his colleagues wrote in their paper. "Two algorithms of interest are the Elastic-Net Attack on Deep Neural Networks (EAD) and TextAttack."

EAD and TextAttack are two algorithms that have proved to be particularly good at creating tampered data that would be misclassified by DNNs. Mathews and his colleagues thus developed a technique for detecting DDOS DNS amplification attacks and then tried to fool it using adversarial data generated by the EAD and TextAttack algorithms.

"In our experiment the EAD and TextAttack algorithms are applied to a Domain Name System amplification classifier," the researchers wrote in their paper. "The algorithms are used to generate malicious DDoS adversarial examples to then feed as inputs to the network intrusion detection systems neural network to classify as valid traffic."

In their tests, Mathews and his colleagues found that the adversarial data generated by EAD and TextAttack could fool their DNN for DDoS DNS amplification attack detection 100% and 67.63% of the time, respectively. These results thus highlight the significant flaws and vulnerabilities of existing deep learning-based methods for detecting these attacks.

"We show that both image processing and natural language processing adversarial learning algorithms can be applied against a network intrusion detection neural network," the researchers wrote in their paper.

In the future, the work by this team of researchers at the Citadel could inspire the development of more effective tools for detecting DDoS DNS amplification attacks, which can detect adversarial data and correctly classify it. In their next studies, the researchers plan to test the effectiveness of adversarial attacks on a particular type of algorithms for detecting DNS amplification attacks, those targeting the so-called

constrained application protocol (CoAP) used by many IoT devices.

**More information:** Jared Mathews, Prosenjit Chatterjee, Shankar Banik, Cory Nance, A deep learning approach to create DNS amplification attacks. arXiv:2206.14346v1 [cs.CR], arxiv.org/abs/2206.14346