AI Based Directory Discovery Attack and Prevention of the Medical Systems

Ying He¹, Cunjin Luo² ³, Jiyuan Zheng¹, Kuanquan Wang⁴, Henggui Zhang⁵

¹University of Nottingham, Nottingham, UK
²University of Essex, Colchester, UK
³Southwest Medical University, Luzhou, China
⁴Harbin Institute of Technology, Harbin, China
⁵The University of Manchester, Manchester, UK

Abstract

The medical system has been targeted by the cyber attackers, who aim to bring down the health security critical infrastructure. This research is motivated by the recent cyber-attacks happened during COVID 19 pandemics which resulted in the compromise of the diagnosis results. This study was carried to demonstrate how the medical systems can be penetrated using AI-based Directory Discovery Attack and present security solutions to counteract such attacks. We then followed the NIST (National Institute of Standards and Technology) ethical hacking methodology to launch the AI-based Directory Discovery Attack. We were able to successfully penetrate the system and gain access to the core of the medical directories. We then proposed a series of security solutions to prevent such cyber-attacks.

1. Introduction

Medical systems have renovated healthcare industry. It has been used to improve the quality of healthcare services by enabling early disease detection and diagnosis which is crucial to the success rate of treatment. Other auxiliary medical services such as personalised medicine also have impact on health and wellbeing. However, the increasing use of technology, fast-increasing health data aggregation and the need of intelligent medical systems in the healthcare industry also led to a series of cyber security issues. It is vital to protect the medical data since compromised data may lead to wrong diagnosis results.

The NHS has experienced and is a victim of the high profile ransomware attack, WannaCry, which has affected a large number of organisations around the globe. After the attack, the healthcare organisations started taking security actions to defend against cyber-attacks [1]. This situation has been compounded by the COVID 19, which further challenged the security of medical systems. Healthcare organisations such as the US Department of Health and Human Services, Brno University Hyjou hospital in Czech Republic, the World Health Organization (WHO) and its Partners have suffered from cyber-attacks during COVID 19 pandemics [2]. It is imperative to protect information in medical systems [3-8].

Ethical hacking which is also called pen-testing can help identify the weaknesses of the system and demonstrate security breaches. It is usually performed following a systematic methodology such as NIST or OWASP in a secure environment. There is existing work demonstrating cyber-attacks towards medical system and presenting security protection solutions [9-14].

In a previous study, we have launched brute force and dictionary attack [14,15]. We have successfully performed the dictionary attack and identified a vulnerability related to authentication “A2 2017-Broken Authentication”, one of the listed OWASP Top 10 vulnerability. In another study, we proposed three attacking activities where AI algorithms are applicable, including password cracking, face recognition and directory discovery [14].

This paper builds on and extends our previous work by proposing a trained algorithm to directory discovery, which is an important type of attacks in ethical hacking.

2. Related Work

2.1. Healthcare Cyber Security

Medical information can be divided into two different types, sensitive and non-sensitive. The information that is relevant to patients such as diagnosis information is classified as sensitive information. For example, the Electrocardiogram (ECG) waveform data and is classified as sensitive information. Such information has been targeted by attackers using various attack vectors. Previous research investigated into ethical hacking in medical system [14, 15], however, AI based ethical hacking has not been well studied in healthcare.

2.2. AI Based Ethical Hacking

AI algorithms can help automate the traditional ethical hacking. The key advantage of AI based ethical hacking are more accurate, can be deployed on multiple systems at a time, it does not reply on human experts, and it is more
suitable for repeatable tests [16].

3. Methods

3.1. NIST Pen-Testing Methodology

The ethical hacking carried out adopted the NIST pen-testing methodology. A complete ethical hacking covers four stages including information gathering, discovery, attack and reporting. [16]. Figure 1 below describes the work flow of NIST pen-testing methodology.

![Figure 1. NIST Pen-testing Methodology [16]](image)

3.2. The Simulated Medical System

To set up the simulated environment, VirtualBox is used as the virtualization tool. We set up a simulation system on the virtual platform through implementing an open-source medical system, OpenEMR. The context of the study is based upon this simulation platform.

![Figure 2. Global view of the algorithm](image)

4. AI based Ethical Hacking

In a previous study, we have launched brute force and dictionary attack [14, 15]. We have successfully performed dictionary attack and identified a vulnerability related to authentication “A2 2017-Broken Authentication”, which is one of the listed OWASP Top 10 vulnerability. In another study, we proposed three attacking activities where AI algorithms are applicable, including password cracking, face recognition and directory discovery [15].

This section proposes applying a trained algorithm to directory discovery.

Directory discovery is an important process in ethical hacking [16]. Spidering can discover directories of web applications by sending HTTP requests and attempting to access other directories by extracting hyperlinks from the response HTML file. Traditional method for directory discovery is the brute force method. In this method, pen-testers create a wordlist and pick URLs in order from the wordlist and then send HTTP requests to the targeted web application. If the response code does not belong to error codes, then the directory is discovered successfully.

This paper proposes a trained algorithm for directory discovery by adopting semantic clustering of sentences. By doing this, directories with the same naming conventions and file extensions are likely to be grouped together so that if one directory within a certain cluster is valid, then other directories within the same cluster are highly likely to be valid. Figure 2 shows the global view of the algorithm.

The wordlist used for directory discovery in this project is composed of absolute paths of directories from three different web applications which are openEMR, DVWA, and Joomla. Firstly, we carried out data-preprocessing process to transform the collected data into sentences. Then, we transferred these into numerical representation by sentence embedding. Finally, the encoded data is fed into the clustering algorithm to create semantic clusters.
5. **Security Solutions**

An important strategy is to enhance the authentication of openEMR. Currently, openEMR directories can be accessed directly by unauthenticated users which can lead to sensitive information disclosure. He, et al., identified that the openEMR session and cookies management is poorly designed which enables cookie stealing [15]. It is advised that openEMR should enforce a secure authentication and session management mechanism which restrict passwords’ minimum size and complexity; authentication is required for every directory; users’ session should be protected, and session should expire when the browser is closed. For wherever possible, multi-factor authentication is always recommended.

6. **Conclusion and Future Work**

Built on previous research on the ethical hacking of the OpenEMR system, this study presented AI-based directory discovery attack and its cyber defence solutions. It has implications for further research into the AI based cyber security attacks and its defence in healthcare. Since the OpenEMR has been widely used in health organisations worldwide, the findings have implications for security professionals in healthcare to prevent against AI based security attacks.

Future work will consider integrating the simulation platform with an intelligent cardiac diagnosis systems with complex computational models such as the arrhythmia detection and classification in ECG data [21-24] and see how the core AI component of diagnosis models can be penetrated and protected. Future work will also consider adopting novel security solutions from other industries to counteract cyber-attacks in healthcare [25-30].

**Acknowledgments**

This work was supported by the National Natural Science Foundation of China (No. 61803318).

**References**


[19] Wang, Ding, and Ping Wang. "Two birds with one stone:


