2022 4th International Conference on Computer Communication and the Internet

Analyzing SocialArks Data Leak - A Brute Force Web Login Attack

Jun Qian, Zijie Gan, Jie Zhang, Suman Bhunia
Department of Computer Science and Software Engineering, Miami University, Oxford, Ohio, USA 45056
Email: qianj4@miamioh.edu, ganz2@miamioh.edu, zhangj63@miamioh.edu, bhunias@miamioh.edu

2022 4th International Conference on Computer Communication and the Internet (ICCCI) | 978-1-6654-6992-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCCI55554.2022.9850265

***Abstract*—In this work, we discuss data breaches based on the "2012 SocialArks data breach" case study. Data leakage refers to the security violations of unauthorized individuals copying, transmitting, viewing, stealing, or using sensitive, protected, or confidential data. Data leakage is becoming more and more serious, for those traditional information security protection methods like anti-virus software, intrusion detection, and firewalls have been becoming more and more challenging to deal with independently. Nevertheless, fortunately, new IT technologies are rapidly changing and challenging traditional security laws and provide new opportunities to develop the information security market. The SocialArks data breach was caused by a misconfiguration of ElasticSearch Database owned by SocialArks, owned by "Tencent." The attack methodology is classic, and five common Elasticsearch mistakes discussed the possibilities of those leakages. The defense solution focuses on how to optimize the Elasticsearch server. Furthermore, the ElasticSearch database's open-source identity also causes many ethical problems, which means that anyone can download and install it for free, and they can install it almost anywhere. Some companies download it and install it on their internal servers, while others download and install it in the cloud (on any provider they want). There are also cloud service companies that provide hosted versions of Elasticsearch, which means they host and manage Elasticsearch clusters for their customers, such as Company Tencent.**

***Index Terms*—Data Breach, SocialArks, ElasticSearch**

I. Introduction

Threat Post published an article in January 2021 that, due to SocialArks' cloud configuration mistakes, 318 million records collected by Facebook, Instagram, and LinkedIn were exposed on the Internet. According to the information from Dotan Nahum. More than 400 gigabytes of social media data and more than 214 million users' information worldwide are involved [1]. This attack is in continuation with the massive data breached discovered over the last couple of years such as the supply chain attack on SolarWinds, Canva data breach, Pegesus spyware data leak [2]–[6].

According to researchers from Safety Detectives, the leak found originated from a Chinese company, SocialArks. During a routine IP address check of a potentially insecure database, the researchers discovered that the server had publicly exposed its records without password protection or encryption [7]. All of the information on SocialArks's server is scraped from the social media platforms such as Snapchat, Instagram, Linkedln, etc., and these platforms do not allow this kind of action. Therefore, scraping information is also a violation of their user agreement. Based on the report by Digital Information World, "there are more than 11 millions user information were from Instagram, more than 66 million of data came from LinkedIn, more than 81 millions of profiles were from Facebook, all of this information was deleted a minute later after the incident came to light." The information leaked includes users' profits, phone numbers, email addresses, hashtags, etc. People can efficiently address a specific person's information from this database [8].

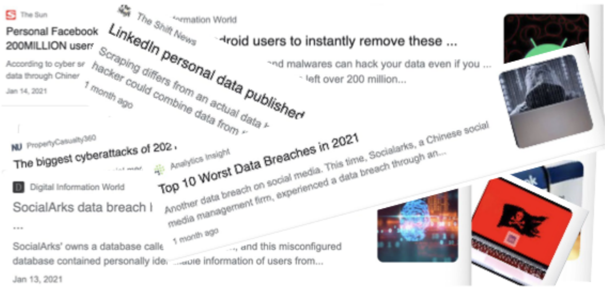
The Sun
Personal Facebook 200MILLION user
According to cyber s data through Chine
Jan 14, 2021
The Shift News
LinkedIn personal data published
Scraping differs from an actual data
hacker could combine data from
1 month ago
rmation World
droid users to instantly remove these ...
nd malwares can hack your data even if you ...
left over 200 million...
PropertyCasualty360
The biggest cyberattacks of 202
Analytics Insight
Top 10 Worst Data Breaches in 2021
Another data breach on social media. This time, Socialarks, a Chinese social media management firm, experienced a data breach through an...
1 month ago
Digital Information World
SocialArks data breach
...
SocialArks' owns a database calle , and this misconfigured database contained personally ide able information of users from...
Jan 13, 2021

Fig. 1. Newspaper headlines right after the data breach [9] [10] [11].

In addition, the database leaked stored personal data as private phone numbers and email addresses that were not open to the public. Nobody knows how they can get the data from these social media. Another point we were noted is that although SocialArks company is based in China, their server is located in Hong Kong where is much away from their company [12]. The company was told about the server vulnerability once the cyber security team discovered it. However, the company did not respond to anything but did secure the server on the same day. Data scraping is a method of extracting private information from social media. It has been common in our daily life due to the rapid spread of seamlessly connected online services. If user authorization is part of consent terms, the data capture is legal given the value of the information obtained. Most data collection is innocuous by web developers, business intelligence analysts, honest businesses, and online marketplaces. However, even if the data is legally obtained, it could lead to a massive leak of information affecting millions of people if it is stored without adequate cyber security.

Moreover, if the data collection is carried out ethically, it will be considered legal. Since the company's servers were