Ahmed Kargbo

CYSE 270

Pro. Yavary

Topic Prompt: Cybersecurity in Education

Cyber and the Social science

Email Scraping

Issue At Hand

Recently several universities including Old Dominion have been receiving many phishing emails and these emails usually specifically target students of the institution. These emails typically contain job offers,raffles, and “database cleanup” for students and it prompts them to reply with personal information in order to secure the job offer, win the raffle draw prize, or provide their organization email and password to perform a database cleanup. I will be explaining the unique way that these private emails are found and also suggesting tactics that organizations can implement to prevent these phishing attempts from being impactful.

What is Email Scraping?

Email scraping is the procedure of using automated tools to collect email addresses from websites, social media sites, or any other online sources. Bots or ‘scrapers’ are commonly used to fulfill this task but email scraping can also be done by a human though it was a lot of time. Nowadays, many scrapers target institutions such as universities or organizations such as corporate jobs and usually pretend to be

someone else that is a part of the said organization or institution eg: CEO, IT HELP. The purpose of email scraping is like many other cyberattacks; to cause spam, financial gain, extract personal data and information etc.

How Does It Work?

The most common and efficient way to scrape emails is by getting an application programming interface (API) that has been programmed to scrape websites. Select a site you would like to engage with and then commence the scraping process. The work is usually done in a three step process:

1. Crawling phase: The programed bot or person starts by navigating through web pages, and follows web links to discover new pages
2. Parsing phase: The bot or person then enters the parsing phase which is simply just analyzing the HTML content. Parsing is a skill used to transcribe syntax and extract information from looking at patterns. After analyzing patterns and uncovering the syntax the bot / person then tests out an example such as; xxvb@gmail.com etc.
3. Extracting: Extraction is the last step. After testing out patterns, lists of email addresses that match the pattern then start to become identified. Once the emails are discovered, they are then extracted and stored in a file.

How to Prevent Email Scraping?

CAPTCHA: Since most of email scraping is done by programgramed bots, it'll be effective for all databases and websites to implement CAPTCHAs. A CAPTCHA which stands for Completely Automated Public Turing test to tell Computers and Humans

Apart, is a security tool used to differentiate between humans and bots thus making it harder for automated tools to gain access and scrape email addresses.

Obfuscation: Obfuscation is a technique that would be beneficial in this event because it'll cause confusion for bots. For example if a bot is trying to scrape an email you can simply change the format to make it harder to understand and duplicate.

EX: xxvb {at} jack {dot} com rather than xxvb@jack.com , this way even if the bot duplicates the format the system wouldn't provide any emails since it'll be unrecognizable.

Honeypots: A honeypot is a system designated to attract cybercriminals and would be highly effective in this situation. By placing hidden fields that aren't visible to users but can be detected by bots is an easy way to detect bots. If a bot fills out one of these fields then the systems can identify and restrict its access.

White Hat Hackers: Many big institutions and organizations typically hire hackers to test their systems for exploitations and vulnerabilities. After discovering vulnerabilities in a system these hackers report them back to said organization and a plan developed. This is a simple yet effective measure but most organizations usually cheap out on this tactic due to 'budget issues'.

Education: Lastly, if all else fails or an organization doesn't implement any of these tactics at all, it is always best to educate users on best and safe cyber practices. Teach users to never click on links listed in emails, double check and verify email addresses, report spam, block unrecognized emails etc.

Conclusion

Cybercrime is continuously on the rise and criminals are continuously getting more and more creative. Though email scraping isn't completely illegal and only considered a crime if used for negative purposes such as obtaining information, spam etc it is still a risky objective. Email scraping is to be used by companies for marketing purposes, but when obtaining emails from college students and targeting them knowing they are highly susceptible, that is going against gray areas and violating laws and is a cybercrime.

References

What is a honeypot and how does it work?

How does security Obfuscation work? | Security Encyclopedia

How CAPTCHAs work | What does CAPTCHA mean? | Cloudflare

What is Parsing? | ITOps Glossary

What is a web crawler? | How web spiders work | Cloudflare

How to scrape emails from any website | ScrapingBee

What is email scraping? How to detect & stop email scraping?