Understanding the Role of AI And Large Language Models in Catfishing on Dating Sites: A Multi-Dimensional Study

Thanh An Vu
Faculty Mentor: Alexi Brooks
University of Wisconsin-Stout

INTRODUCTION

The definition of online dating deception, related to the term “catfishing,” involves approaching a potential romantic partner online without genuine intentions of creating an in-person relationship.

This study aims to demonstrate algorithms for detecting fake profiles on online dating sites based on a provided dataset. Current online dating sites already have their security layers. However, they only work for low-sophistication attacks, not higher ones. I present two vulnerability concepts that may increase the fallibility of a detection system.

APPROACH

I explain the security techniques that online dating platforms are applying to protect users**.** Utilizing the idea of a large language model to detect online catfishing and provide related algorithms to help the investigation.

Many online dating platforms on the market are currently in terms of target audience, appearance design, functions, and popularity in today's world. Followed by Forbes, some online dating services attracted clients such as Tinder, Bumble, Elite Singles, Silver Singles, Plenty of Fish, and Match. These applications let users very their preferences when deciding on an attractive partner whom they want to have a conversation with. The suggestions from online dating sites will recommend a person based on common interests, goals, or personal descriptions that clients put in. In addition, the system will consider a “match” when both users give each other a “like”.

In the past, dating websites already have some techniques for detecting fake accounts. For example, Jiayuan (a popular dating platform in China) has a behavioral-based, IP address-based, photograph-based, and text-based detection system to recognize fake accounts. Specifically, scammer accounts will be flagged if they use the same IP address to create multiple accounts, using duplicating images or similar messages to potential partners. However, these techniques only work for low-sophistication attacks when scammers have little understanding of the technology.

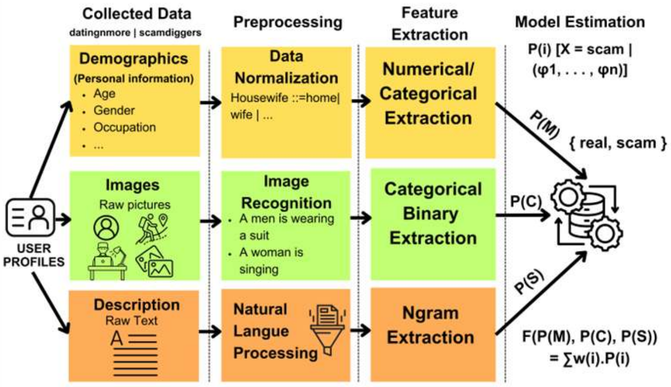
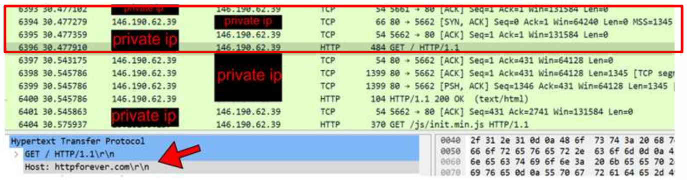


Figure 1: Machine learning model based on Suarez-Tangil et al (2020).

- For high-sophistication attacks in which most of the profiles are real, the traits are more varied. Those listed above in Figure 1 will work for most available online dating sites, including: demographics, images, and descriptions. Each of the classifiers deals with missing data because those fields are not mandatory for users to complete due to the private preferences policy before calculating the probabilities of θ_M, θ_C, and θ_S. Using only one classifier will lead to less accurate results instead of applying multiple ones.

Potential Vulnerabilities:

- **New Users Never Use Dating Services**: There exists the possibility of creating an account on dating sites based on a real profile on other social media accounts that have not used online dating sites before. Hence, increasing the possibility of wrong predictions of a fraudulent account.
- **Replayed Attack:** This is a technique in which hackers can intercept and retransmit data that was recorded in the system by using hacking tools. This can happen when people open online dating platforms while using public free Wi-Fi. Attackers can capture the data transmitted including authentication credentials or encrypted messages and then send it back to trick the server. By taking advantage of the Three-Ways Handshake when a device accesses a website, attackers can take over user account.



```
6393 30.477102  private ip      146.190.62.39  TCP     54 5661 → 80 [ACK] Seq=1 Ack=1 Win=131584 Len=0
6394 30.477279  146.190.62.39   private ip     TCP     66 80 → 5662 [SYN, ACK] Seq=0 Ack=1 Win=64240 Len=0 MSS=1345
6395 30.477359  private ip      146.190.62.39  TCP     54 5662 → 80 [ACK] Seq=1 Ack=1 Win=131584 Len=0
6396 30.477910                  146.190.62.39  HTTP   484 GET / HTTP/1.1
6397 30.543175  146.190.62.39                  TCP     54 80 → 5662 [ACK] Seq=1 Ack=431 Win=64128 Len=0
6398 30.545786  146.190.62.39   private ip     TCP   1399 80 → 5662 [ACK] Seq=1 Ack=431 Win=64128 Len=1345 [TCP segr
6399 30.545786  146.190.62.39                  TCP   1399 80 → 5662 [PSH, ACK] Seq=1346 Ack=431 Win=64128 Len=1345 |
6400 30.545786  146.190.62.39                  HTTP   104 HTTP/1.1 200 OK  (text/html)
6401 30.545863  private ip      146.190.62.39  TCP     54 5662 → 80 [ACK] Seq=431 Ack=2741 Win=131584 Len=0
6404 30.575937                  146.190.62.39  HTTP   370 GET /js/init.min.js HTTP/1.1

Hypertext Transfer Protocol              0040  2f 31 2e 31 0d 0a 48 6f  73 74 3a 20 68 7
 > GET / HTTP/1.1\r\n                    0050  66 6f 72 65 76 65 72 2e  63 6f 6d 0d 0a 4
   Host: httpforever.com\r\n             0060  6e 65 63 74 69 6f 6e 3a  20 6b 65 65 70 2
                                         0070  69 76 65 0d 0a 55 70 67  72 61 64 65 2d 4
```

Figure 2: Using Wireshark to capture Three Ways Handshake Packets when accessing a website using HTTP protocol: httpforever.com

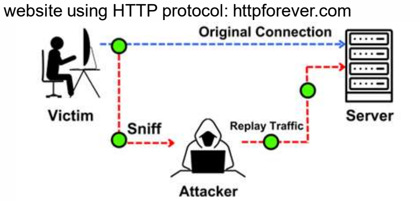


Figure 3: Replay attack simulation

- **File Upload Vulnerability:** This vulnerability may affect the system because all users are required to upload at least one profile picture on a dating service. Most web-based applications are written in JavaScript or PHP languages and the format of popular uploaded pictures (JPEG or PNG). Attackers can utilize this to upload other web files such as (test.php) or (test.js.png/ test.php.jpeg) to access the local data.

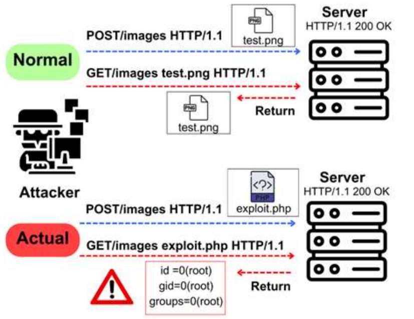


Figure 4: File upload vulnerability simulation

DISCUSSION

- In terms of using AI algorithms to detect fake accounts on dating sites, we also have to mention the importance of the probability of the algorithms for misclassifying known as True vs. False and Positive vs. Negative classifications as fake profiles can be detected as real ones.
- I believe the vulnerabilities I mentioned would affect user experience and reputation of dating service companies. Especially in the case where scammers can steal people’s profile information on a specific social media app because we tend to share more information online currently.
- It is ethical to find the vulnerabilities of a system to improve the user’s experiences and prevent security threats. Additionally, I did not create an account to avoid the situation that my profile would become a potential partner for real users on dating sites and use it for illegal purposes.

	Scam	Real
Scam	**True Positive (TP):** Scam profiles correctly classified as scams.	**False Positive (FP):** Real profiles misclassified as scams
Real	**False Negative (FN):** Scam profiles misclassified as real.	**True Negative (TN):** Real profiles correctly classified as real.

Figure 5: True-False Positive and Negative Prediction Model

REFERENCES

Suarez-Tangil, G., Edwards, M., Peersman, C., Stringhini, G., Rashid, A., & Whitty, M. (2020). Automatically Dismantling Online Dating Fraud. *IEEE Transactions on Information Forensics and Security*, *15*, 1128–1137. https://doi.org/10.1109/TIFS.2019.2930479

Huang, J., Stringhini, G., & Yong, P. (2015). Quit playing games with my heart: Understanding online dating scams. *Detection of Intrusions and Malware, and Vulnerability Assessment*, 216–236. https://doi.org/10.1007/978-3-319-20550-2_12