BaitBlock: Measuring and Mitigating Phishing Propagation in YouTube Live Chats

Andy Wu
cw4483@nyu.edu
B.S. Computer Engineering

1 Introduction

1.1 Problem Statement

Live streaming platforms like Twitch and YouTube Live have become an incredibly effective medium for novel phishing and impersonation scams. Attackers piggyback on the trust and parasocial relationships of communities surrounding different streamers to develop scams. Recent research reveals the alarming scope of this issue. A joint study between UCSD and Google [2] measuring cryptocurrency giveaway scams found that scammers converted about 4 in 100,000 live stream views into victims, extracting nearly $4.62 million from just a few hundred people during the study window.

1.2 Motivation and Significance

Notably, younger fans are especially susceptible to this new form of phishing attack. While children are already more vulnerable to scams, streaming platforms have a predominantly young audience and exacerbate children's vulnerability by providing a space for attackers to target them easily. A young audience who are eager to interact with their favorite streamers are much more easily fooled by attackers impersonating the same streamers. A multi-year study from RiskIQ [1] revealed how criminals mimicked seven prominent YouTube channels, such as vlogger James Charles and commentator Philip DeFranco, and successfully tricked over 70,000 viewers by sending messages that framed malicious links as links to receive giveaway prizes.

Live streaming as a platform is relatively young, and academic attention has lagged behind more established social media. Phishing has been studied extensively in emails and static social network posts, but YouTube Live and Twitch chats present a new terrain that is fast-moving and loosely moderated. It is an incredibly difficult challenge to prevent this new form of phishing attack entirely. Streamers do frequently organize authentic promotional events like giveaways and raffles that are incentivized by cryptocurrency. They reach out to fans and winners in the same communication channels (e.g, direct messages, stream chatrooms) where impersonators attempt scams. Occasionally, the official events sponsored by prominent streamers are scams themselves. Numerous high-profile NFT and meme coin projects promoted by very popular streamers have ended in pump-and-dump collapses, with a majority of their participants suffering serious financial losses. This further blurs the boundaries between legitimate and malicious engagements on streaming platforms. Consequently, it is difficult to build effective rule-based filter solutions against phishing attacks on streaming platforms.

1.3 Scope and Project

BaitBlock will be scoped as a platform measurement project that will measure and flag impersonation-driven phishing scams. Due to time restrictions, BaitBlock will focus on monitoring streams on YouTube Live only. The tool under development is a Chrome Extension that parses any YouTube live chatroom in real-time, flags high-risk scam content, and cross-references it with platform moderation events. I will be focusing specifically on streams categorized under "Finance," "Web3," and "Influencer". These three genres are particularly saturated with scam behavior, according to both anecdotal and empirical studies.

1.4 Project Outcomes

2 Related Work

2.1 Core Sources

2.2 Identify Gaps

3 Research Plan and Current Status

3.1 Objectives

3.2 Methodological Approach

3.3 Progress to Date

3.4 Next Steps

References

[1] Anthony Cuthbertson. Youtube impersonation scam has tricked 70,000 people, study reveals. *The Independent*, January 2019. Accessed: 2025-11-02.

[2] Enze Liu, George Kappos, Eric Mugnier, Luca Invernizzi, Stefan Savage, David Tao, Kurt Thomas, Geoffrey M. Voelker, and Sarah Meiklejohn. Give and take: An end-to-end investigation of giveaway scam conversion rates. In *Proceedings of the 2024 ACM on Internet Measurement Conference (IMC '24)*, pages 704–712, 2024. arXiv preprint arXiv:2405.09757v1.