**Privacy: Lessons From Your Local Drug Dealer**

*Karl Eirich, Kasha Muzila*

*UC Berkeley MIDS and MICS, W233, Summer 2022*

### **1. Introduction: Why is the problem relevant or important?**

Making purchases online and exchanging money with friends has never been easier with apps such as Venmo. Especially with the exponential growth in businesses transitioning to cashless transactions fueled by the COVID-19 pandemic to avoid physical touch in making purchases [1]. This growth came with new datasets; some were available to the public and could contain sensitive information that the users of these apps may not want exposed.

Venmo, in particular, poses a unique problem due to the app's belief that it is also a social network [2]. This can expose sensitive information such as where they live, what school they go to, their daily schedule, their network of friends, and even illegal or controversial activities. With over 60 million people using the app [3], users can see each other's purchases by default, unbeknownst that their sensitive information is public. This can give an attacker everything they need to know about their victim without much effort.

With nearly a third of millennials using Venmo to pay for drugs [4], drug dealers had to develop unique ways to protect their network and privacy to avoid getting caught. An example would be using a continuously evolving emoji language [5] to create ambiguity in their transactions. This can make it substantially more difficult for an attacker to access their sensitive information and their network.

To investigate how much more private a drug dealer that uses Venmo versus an average user, we accessed a public Venmo database created by Github user ***sa7mon***. This dataset has over seven million records scraped from the Venmo public API between mid-2018 to early 2019 [6]. The Github user made this dataset available to bring attention to Venmo users that all of their data is publicly available for anyone, including attackers, to grab without an API key [6].

Our project aims to evaluate Venmo's transaction threat to user privacy and identify methods to preserve privacy by studying drug dealers' Venmo data. The results of our project are based on exploratory data analysis and research into drug dealer behavior within social media to give Venmo and its users better privacy practices to help mitigate potential attackers.

### **2. Background and Survey of related work**

Some 150 million American adults say they have swapped cash and credit for digital wallets. About 66% of those surveyed say they use digital wallets because they are more convenient than carrying around dollar bills and credit cards [7]. However, only about 9% of adults read the app's privacy policy in its entirety before agreeing to the terms and conditions [8]. This can cause app users to be at risk of companies jeopardizing app users' online privacy for attackers to use against them.

One study done by Rajat Tandon titled *I know what you did on Venmo: Discovering privacy leaks in mobile social payments* [9] suggested that Venmo's public-by-default policies can put users at risk of criminal investigation, theft, health benefit loss, job and opportunity loss, national security risks, identity theft and financial scams, and adverse emotional effects. They found that 41 million notes (10.5%) leak sensitive information such as health condition, political orientation, and drug/alcohol consumption involving 8.5 million (37.8%) users over eight years of data. They also found that a user who transacts with a group on Venmo can be reduced or eliminated through the actions of other users. This behavior typically occurred around half of Alcoholics Anonymous, gambling, and biker gang members.

Another study done by Xin Yao titled *Beware of What You Share: Inferring User Locations in Venmo* [10] used multilayer location inference (MLLI) techniques to infer user locations for the public transaction records made available by Venmo. Their MLLI explored two observations: the first was that many transaction notes contained implicit location cues, and the second was that user transactions' types and temporal patterns have strong ties to their location closeness. After analyzing 2.12 million users and 20.23 million Venmo transaction records, they showed that their MLLI could identify the top-1, top-3, and top-5 possible locations for a Venmo user with an accuracy up to 50%, 80%, and 90%, respectively.

Previous research by Dan Gorelick, the Github user ***sa7mon*** that created the public Venmo database, showed how easy it was to scrape Venmo's public data [11]. This seven million record dataset scraped from the Venmo public API spanning from July 2018 to February 2019 was made public on Github to bring attention to Venmo users that all of this data is publicly available for anyone to grab without using an API key. He then states that there are some precious data for any attacker conducting open source intelligence (OSINT) research and recommends users switch their Venmo account to private.

There have been previous W233 projects that have researched this Venmo dataset. However, our study uses the previously mentioned work as background for Venmo's concerning privacy policies. Our study expands on how drug dealers could adopt better privacy strategies to safeguard their sensitive information while using Venmo versus the average user and how Venmo can adopt more secure privacy policies based on our observations.

**2.1 Privacy Risks**

While Tandon's study was more specific in the types of risks a Venmo user can experience, we conducted our own risk assessment regarding what we learned in W233. Those risks are Identity Disclosure, Attribute Disclosure, Membership Disclosure, and Inference risks. The following descriptions are from analyzing the Venmo dataset individual transaction records and its categorical (column) labeling.

*Identity Disclosure* – The definition of this risk is being able to tell the identity of the person to whom the record corresponds (anonymity) [12]. All the individual transaction records within the Venmo data are composed of the sender and receiver's usernames, first and last names, and pictures of what they could look like. Our research found that if a Venmo user links their Facebook account, their Facebook ID is also listed within the dataset.

*Attribute Disclosure* – The definition of this risk is being able to tell that a person has a specific (sensitive) attribute [12]. The information within the Venmo dataset that would classify as being within attribute disclosure are the sender and receiver's email, phone numbers, if the transaction was created and completed, and whether they made the transaction using Android or iPhone.

*Membership Disclosure* – The definition of this risk is being able to tell that a person is in (or not in) a dataset (confidentiality) [12]. Due to the amount of information given out within the identity and attribute disclosure risks, finding a specific person within the Venmo dataset is effortless. The studies mentioned previously in our survey of related work also mention how effortless it is to find someone within the Venmo dataset.

*Inference Threat* – The definition of this risk is being able to tell something new (undisclosed) about a person [12]. The information within the Venmo dataset that would be classified under this category are whether they are friends with each other, whom they are paying, when they joined Venmo, their "About Me" section of their Venmo profile, if it was labeled as a "payment" or "request", and the transaction note that describes what they purchased. For clarification, "payment" refers to the sender sending a payment to the receiver, and "request" refers to the receiver requesting a payment from the sender. The transaction note can also be classified as an attribute disclosure threat depending on what was written within the note.

### **3. Methods: What did we do and how we did it**

The ingestion of the dataset required some planning to ensure the active service of our analysis objective with minimal maintenance cost. Our planning resulted in creating a local MongoDB server to store the dataset and some code snippets to facilitate connecting our Jupyter notebook to the server. While implementing our dataset ingestion plan, we familiarize ourselves with the Venmo API transaction schema to establish our analysis scope and identify fields of interest. The three fields that were primarily used during our analysis were “payment.actor.username” (buyer), “payment.target.user.username” (seller), and “note” (description). Focusing on these fields allowed for quick identification of drug deals and easy pivoting to examine a potential drug dealer.

We decided to set the project's scope of analysis to "payment" type transactions with drug-related emojis in the "Note" field. The decision to focus on drug-related emojis, but not drug-related terms, stems from Venmo's implementation of a banned word list that includes drug-related terms preventing their use in a transactions note [13]. Limiting the transaction type to "payment" allows us to assume the target (recipient) of a drug-related transaction is the drug dealer.

Emojis carry a level of implicit ambiguity due to their function as a pictogram embedded in the text to convey emotional cues [14]. The adoption of specific emojis as slang for drugs can add additional ambiguity to those emojis when trying to determine what they are communicating. This forced us to rely on analyst inference rather than an established ground truth or statistical methods. The varying interpretations of a specific emoji meant that there would be False Positive detections, and we would need to use our informed intuition to determine their acceptability.

Our initial filter to surface drug deals in the dataset was based on the Drug Enforcement Administration (DEA) "Emoji Drug Code" document that mapped emojis to the drug it represented [15]. The DEA document is heavily context-dependent, but the results of the initial filter were promising. However, we had to come up with some of our own interpretations due to some emojis being used too frequently in everyday transactions. The ambiguity around an emoji's intended interpretation grew exponentially, preventing us from using it in our filter. The ambiguity's growth was partly fueled by the dataset's significant imbalance between normal and illicit behavior. We did expect an imbalance going into the analysis. However, the surge of false positive detections would need to be mitigated with every addition to the filter, creating doubt around our interpretation of the emoji. We settled the dataset when the drug deal filter results had an acceptable precision/recall balance, and new emojis showed diminishing returns.

The emojis in Table 1 represent opposite ends, with the left column containing highly interpretable emojis and the right column holding highly ambiguous emojis. The eightball and nose are both slang for cocaine, with the nose frequently appearing alongside other emojis that provide the necessary context to interpret the meaning of the transaction's note. One example is the nose (👃) and Mt.Fuji (🗻) appearing together to represent cocaine, while the nose (👃) and bear (🐻) together seem unrelated to drugs and could represent something simple like allergy medication. The mushroom emoji (🍄) is slang for magic mushrooms (Psilocybin) and appears alongside other drug-related emojis. An example is "🍁🍄❄️💊🤫😜😈" which also includes playful emoji faces that are literally winking the meaning of the transaction note. The pill emoji (💊) is slang for pill-based drugs such as MDMA, Molly, Xanax, Percocet, or Adderall, which holds a secondary slang inference as a general representation of drugs in general. The emojis on the right are used to represent a variety of everyday activities done frequently, such as the lightning emoji (⚡) representing their electric bill, the diamond emoji (💎) broadly representing jewelry purchases, and the skier (⛷) represents any snow related activity.

| True Positive Emojis | False Positive Emojis |
| --- | --- |
| 🎱 | ⛷ |
| 👃 | 💎 |
| 🍄 | ⚡ |
| 💊 |  |

**Table 1**

### **4. Results: How it all worked out**

**4.1 Dealer Privacy**

When measuring the privacy of a drug dealer, we initially took a holistic approach looking at transactions where they were both target and actor. We discovered that many of the drug dealers were not the actors in any transaction or the transactions were innocuous. After some debate, we reframed the measure to identify if the average number of public transactions targeting drug dealers is lower average than a typical user. Based on our aggregation, we did find that drug deals received fewer publicly available transactions than the average user, as seen in Table 2.

|  | Normal User | Drug Deal |
| --- | --- | --- |
| Average Number of Transactions Received | 2.85 | 1.04 |
| Total Number of Transaction | 4050457 | 73688 |

**Table 2**

**4.2 Personal Relationships Between Users**

While exploring drug deal transactions, we discovered a swath of Venmo transactions that mainly occurred between a pair of users. It provides a detailed log of information about the pair that would be useful to a motivated individual targeting them. Evaluating the transaction history (Image 1) through the lens of an individual targeting the pair, the immediately noticeable signal is the frequent occurrence of the eight-ball emoji, commonly used as slang for cocaine. The repeat occurrence of the taco emoji and ‘in out’, a potential misspelling of ‘In-N-Out’, a popular fast food restaurant primarily in California, is a solid lead that the pair is located on the West Coast, potentially in California. A transaction with the note ‘Offffwhitesssss’ suggests one of the pair is interested in fashion by their enthusiastic note attached to the purchase of an item by the famous luxury brand ‘Off White’. Finally, the pair mainly interacts with themselves. In contrast, one pair repeatedly interacts with a handful of other users that can be leveraged as potential entry points or additional sources of information to validate assumptions and expand the pool of signals about the pair.

### **5. Discussion: What does it all mean?**

**5.1 Drug Dealer Lessons**

Venmo's belief that it can exist as a social network appears to be at odds with the privacy needs of users who believe they signed up for a payment app. We can see this demonstrated in the origin of our dataset to our ability to make data-backed inferences about users ever leaving Venmo's ecosystem. Drug dealers appear to have the habit of finding and enabling privacy settings as a first step whenever possible is something everyone should adopt. We recommend enabling the privacy settings on Venmo not only to protect their spending habits but also to protect their sensitive personal identifiable information (PII) from being distributed publicly.

**5.2 API Suggestions**

Our team recommends that Venmo considers removing many fields from the public availability transaction data. An example would be the inclusion of a username, display name, first name, last name, and id of both users in the transaction is excessive. Removing users' first and last names would be a big step toward protecting their users' sensitive data. Adding granular data access controls and throttling the speed at which available data can be scraped would also welcome improvements.

**5.3 App Suggestions**

To improve the app, our team recommends increasing the granularity of control over user data, adding more friction or reminders encouraging users to review their privacy settings, and making all newly created user accounts private by default.

**5.4 Privacy/Utility Trade off**

While working on the project, we were consistently amazed at the quantity of data Venmo makes publicly available. We could not find any functional reason for most of the data in a transaction to be shared publicly. As a team, we concluded there would be no loss of utility by removing many of the available fields from the public API. At the same time, the improvements to user privacy would be immense.

### **6. Members' contributions: Who did what**

All members equally contributed to the research problem, different methods of analysis, and conclusions from the results of our study.

*Karl Eirich* – Data ingestion, data processing, data analysis, research into emoji interpretation, report, presentation

*Kasha Muzila* – Data exploration, data analysis, data visualizations, identified privacy threat insights, survey of work, report, presentation

**7. References**

1. https://www.bis.org/statistics/payment\_stats/commentary2112.pdf
2. https://www.cnet.com/news/privacy/venmo-explains-why-transactions-are-public-by-default/
3. https://venmo.com/
4. https://9to5mac.com/2022/05/19/drug-deals-via-apps/
5. https://lendedu.com/blog/nearly-third-millennials-used-venmo-pay-drugs/#:~:text=Nearly%20a%20third%20of%20respondents,app%20to%20pay%20for%20narcotics.
6. https://github.com/sa7mon/venmo-data
7. https://www.finder.com/digital-wallet-statistics
8. https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/
9. https://www.researchgate.net/profile/Rajat-Tandon-2/publication/361101746\_I\_know\_what\_you\_did\_on\_Venmo\_Discovering\_privacy\_leaks\_in\_mobile\_social\_payments/links/629cbeef416ec50bdb0e09e4/I-know-what-you-did-on-Venmo-Discovering-privacy-leaks-in-mobile-social-payments.pdf
10. https://web.asu.edu/sites/default/files/cnsg/files/j38.pdf
11. https://danielgorelick.com/scraping-venmo/
12. W233 Lecture
13. https://www.gawker.com/the-words-that-will-get-you-in-trouble-on-venmo-1738331672
14. https://www.intechopen.com/chapters/69271
15. https://www.wdrb.com/news/crime-reports/dea-unveils-emoji-drug-code-to-educate-parents-about-drug-exchanges/article\_ca22afc6-aed2-11ec-b478-2b19a331d58f.html#:~:text=The%20agency%20shared%20a%20list,is%20code%20for%20all%20drugs.