Address obfuscation

Signal exploration research - Gyorgy Mora, DS Team

The problem

Some of our customers have experienced fraudulent bypassing of address-velocity-based anti-fraud measures by obfuscating the addresses so that the product was still delivered to an accessible location.

Example

A retailer tries to limit purchase quantity per customer of a particular product by enforcing limits in the number of items purchased per shipping address. Fraudsters obfuscate shipping addresses, thereby obtaining product quantities over the limit and resell the product on a secondary market, earning a huge margin. Following is a mock example of address obfuscation:

Original address: *1301 5th Ave #1600, Seattle, WA 98101, United States*

- 1301**a** 5th Ave #1600, Seattle, WA 98101, United States
- 1301**b** 5th Ave #1600, Seattle, WA 98101, United States
- 1301 5th Ave #1600, Seattle, WA**123** 98101, United States
- 1301 5th Ave #1600, Seattle, WA 98101**-abc**, United States

Current capabilities

While some of the obfuscated addresses above will be normalized to the same address by our address normalization, parts of the address will allow obfuscation to result in separate address digests, thereby bypassing checks. Currently our customers use the address digest to enforce policies and to detect high velocity addresses in other types of fraud. Therefore, obfuscated addresses yielding different digests for the same address will allow fraudsters to keep the address velocity low.

While Ekata offers advanced address parsing and normalization, its performance varies by region and no rule-based system can cover all current and future attempts to obfuscate addresses.

Explored solutions

Fraudsters are able to bypass the velocity checks because obfuscated addresses are not always normalized to the same address, and Ekata customers rely on the address digest of the normalized address to measure velocity. We explored solutions that calculate an alternate velocity-like signal which is resilient to any form of obfuscation. The proposed solution is also suitable for addresses in unsupported formats and addresses in countries where we have less coverage.

Character 3-grams

Character 3-grams are all the 3-character long substrings of an address:

```
1301 5th Ave #1600, Seattle, WA 98101, United States ->
   1. 130
   2. 301
   3. 01_
   4. 1_5
   5. _5t
   6. ...
```

The basic logic of using a character-3 gram is that obfuscation will only affect a small percentage of all trigrams in an address, therefore calculating velocities based on the trigrams is likely to eliminate the problem. It will also make normalization obsolete for velocity calculation.

First approach

We created a data-set combining a set of known obfuscated addresses from customers with addresses from non-hashed test harvesteable data as “good” addresses. We calculated statistics on the 3-grams in the addresses in order to find differences between the obfuscated addresses and the control group. By manual inspection of the data we observed that known obfuscated addresses can be separated by the trigram statistics. We even found some obfuscated addresses within the harvesteable test data in our “good” dataset.

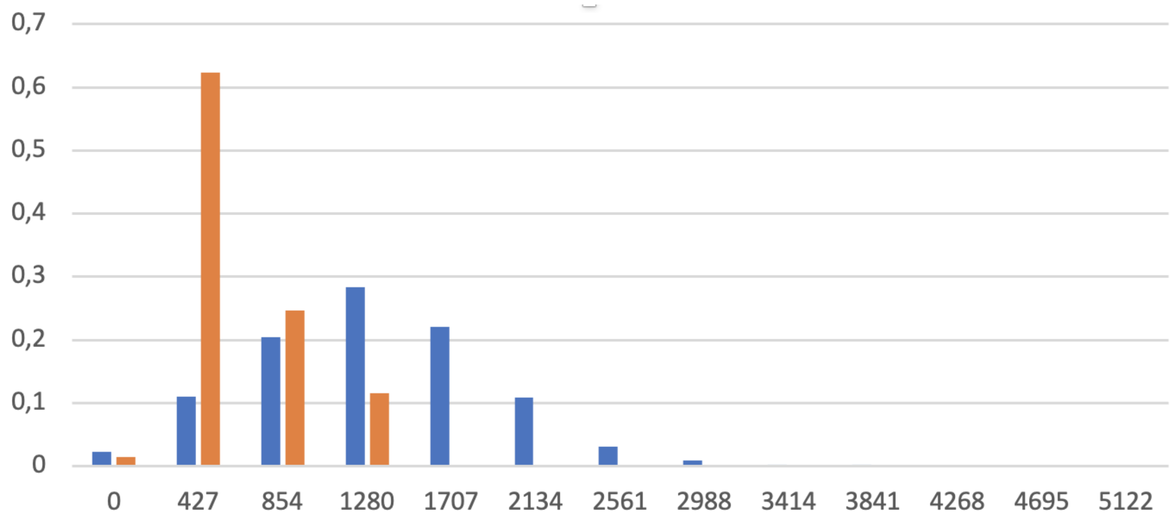

Figure 1: Distribution of the average trigram counts for normal (blue) and obfuscated addresses (orange). Obfuscated trigrams occur with low frequency and shift the distribution to the left.

Second approach

The first approach mistakenly identified a certain kind of address as obfuscated, namely an address in which more than one element contains frequent trigrams : **`1234 Seattle St, Buffalo, New York, USA`** contains frequent trigrams and all addresses from this street might be considered obfuscated.

To overcome this issue, we formed pairs of generated trigrams, and calculated the same statistics as for trigrams over the trigram pairs. This approach avoids depending on two generally frequent trigrams – **`Sea`** from Seattle and **`New`** from New York – and instead uses the less frequent trigram pair **`SeaNew.`**

Trigram pairs are only considered frequent if a specific address containing both trigrams was frequent in the past 24 hours and not if addresses containing just the parts of the trigram pair were frequent separately. This event will be interpreted as possible fraud. Trigrams from the obfuscated parts will likely have low frequency (because they are unique).

It is also important to note that it is not one trigram or one single trigram pair which generates a fraud signal but rather all the trigrams and pairs in the address. For a "normal" address, the counts of the trigrams/trigram-paris will follow a long tail distribution. But for a fraudulent and obfuscated address there will be an almost uniform distribution of high-frequency trigrams. This difference in distributions should be condensed as a signal. Further research should determine the best approach for creating a single signal from the individual statistics.

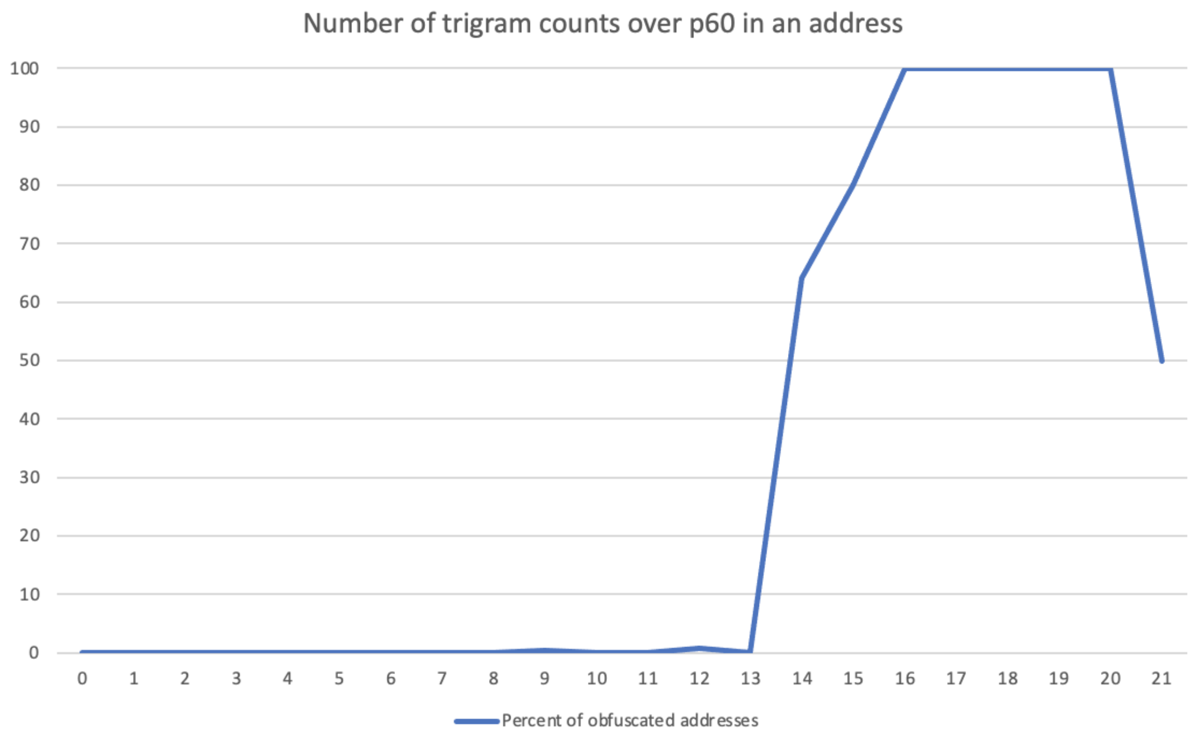

Figure 2: One of many signal candidates. In each address the counts of the trigrams are calculated. A signal results when the number of trigrams in

the address is above p60 counts. We try to catch obfuscated addresses where most of the trigrams have an elevated count due to address replication, not just a few frequent ones. The signal correlates with the percent of addresses being obfuscated.

Conclusions

The trigram pair function showed promising first results, but the dataset we used was small. Therefore further investigation is needed to find the most discriminating signal. Nevertheless, the average count of trigram pairs in the preceding 24 hours demonstrates a separation between obfuscated and good addresses but the exact connection needs to be determined

The data we used contained only a dozen different addresses and slightly more than a hundred transactions. Therefore, we do not consider the result conclusive. Still, the outcome does not contradict the theory that trigrams or trigram pairs can be used to construct effective signals to address the problem of fraudulent address obfuscation.

Next steps

- Tests on synthetic data to find the best statistic to be used as a signal.
- A bigger set of obfuscated addresses (or labeled transactions) is needed to develop actual signals usable on production data.
- Data from multiple customers/industries with different obfuscation methods is needed to ensure we are able to handle all of them.