Synthetic Financial Data

Money Laundering

For use in AML: Anti-Money Laundering

Erik Altman

ealtman@us.ibm.com

February 25, 2021

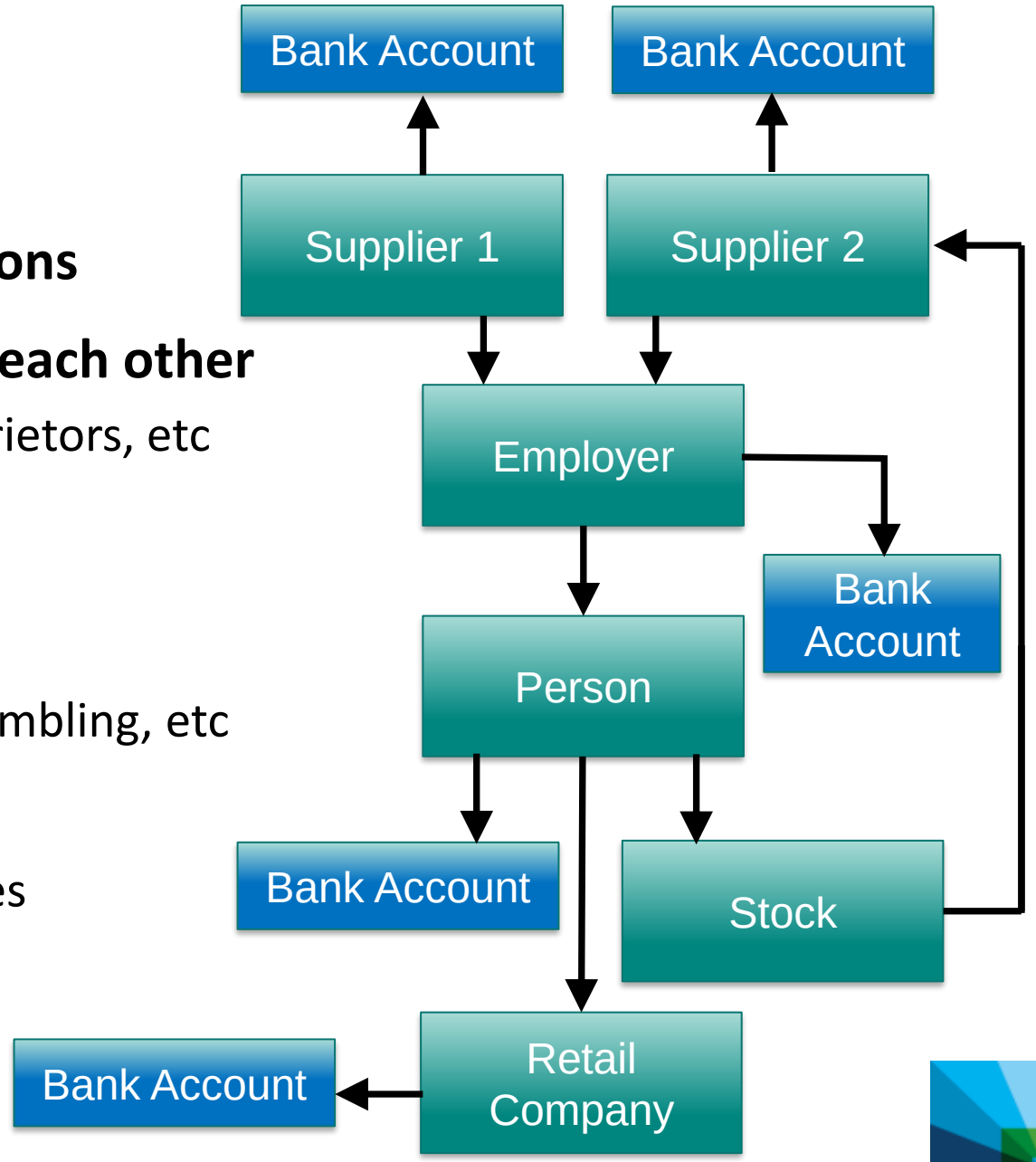
https://github.com/IBM/AML-Data

https://ibm.box.com/v/AML-Anti_Money_Laundering-Data

Bank Transfers: AML – Anti Money Laundering

AML Approach

- **Data generation similar to credit card transactions**
- **Multiple actors in a virtual world interact with each other**
  - Individuals, corporations, partnerships, sole proprietors, etc
  - Cross-linked ownership graph
  - Salaries, pensions, bank interest, dividends, etc
  - Accounts receivable, daily purchases, etc
  - Proceeds of smuggling, extortion, illegal drugs, gambling, etc
- **Complex connections between entities**
  - Unlike credit card data: consumers and companies
  - → Bipartite graph



Reasons to Transfer Money

Legit

- Pay for Goods
- Pay for Services
- Higher Interest Rate
- Monthly Savings
- Avoid Negative Balance

Note: Services can include items such as employee salary

8 Money Laundering Patterns

- Fan-Out
- Fan-In
- Cycle
- Bipartite
- Stack
- Random
- Scatter Gather
- Gather Scatter

Method of Data Generation and "Detecting" Laundering

- With our multi-agent virtual world approach to generating data, there is no need to "detect" laundering -- as laundering is implicit in the actions of the model. That is the underlying model is of individuals and companies. A few of those individuals and companies engage in illicit activity and obtain funds through that illicit activity. When those criminal individuals and companies transfer their illicit funds, the transactions are automatically laundering -- with no need to "detect" them.
- The laundering tag on financial transfers is also transitive. For example, if A pays B $100 in illicit funds (via a ACH, Credit Card, Check, Wire, etc), and B then pays $50 of those funds to C, and C pays $25 of the $50 to D, then not only the initial $100 is laundering, but the $50 and the $25. This transitivity is supported in the underlying model where all funds are tagged with whether their original source is legitimate or illicit.
- The model is also constructed so that some of those laundering transfers happen "naturally", e.g. a criminal boss launders money by paying employees, buying supplies, etc. Other of the laundering transfers are more methodical laundering and follow 1 of the 8 "standard" AML patterns (scatter-gather, cyclic, etc) on the previous page.

AML – Sample Transaction Data

Timestamp	From Bank	Account	To Bank	Account	Amount Received	Receiving Currency	Amount Paid	Payment Currency	Payment Format	Is Laundering
1/1/2019 0:22	800319940	8004ED620	808519790	872ABC810	120.92	US Dollar	120.92	US Dollar	Credit Card	0
1/1/2019 0:05	8021ADE00	80238F220	9A7F59FA0	A23691240	33.97	US Dollar	33.97	US Dollar	Credit Card	1
1/1/2019 0:14	801946100	8023F0980	83585F5A0	948893910	79.20	US Dollar	79.20	US Dollar	Credit Card	0
1/1/2019 0:05	80010C840	800122AA0	80010C840	800122AA0	8,834.09	Euro	10351.64	US Dollar	ACH	0
1/1/2019 0:05	80010C840	800122AA0	80010CF20	80012DA00	8,834.09	Euro	8834.09	Euro	ACH	0
1/1/2019 0:08	80010CF20	80012DA00	80010CF20	80012DA00	9,682.16	US Dollar	8262.75	Euro	ACH	0
1/1/2019 0:08	80010CF20	80012DA00	80010BD60	80011E460	9,682.16	US Dollar	9682.16	US Dollar	ACH	0
1/1/2019 0:03	800319940	800466670	80029A010	8002F6F20	9,125.22	US Dollar	9125.22	US Dollar	ACH	0

Other Currencies

Yuan
Yen
Indian Rupee
Ruble
UK Pound
Canadian Dollar
Australian Dollar
Mexican Peso
Brazilian Real
Swiss Franc
Shekel
Saudi Riyal
Bitcoin

Other Formats

Wire
Cheque
Cash

AML – Sample Transaction Data (CSV File Format)

Timestamp,From Bank,Account,To Bank,Account,Amount Received,Receiving Currency,Amount Paid, Payment Currency,Payment Format,Is Laundering

2019/01/01 00:22,800319940,8004ED620,808519790,872ABC810,120.92,US Dollar,120.92,US Dollar,Credit Card,0

2019/01/01 00:05,8021ADE00,80238F220,9A7F59FA0,A23691240,33.97,US Dollar,33.97,US Dollar,Credit Card,1

2019/01/01 00:14,801946100,8023F0980,83585F5A0,948893910,79.20,US Dollar,79.20,US Dollar,Credit Card,0

2019/01/01 00:05,80010C840,800122AA0,80010C840,800122AA0,8834.09,Euro,10351.64,US Dollar,ACH,0

2019/01/01 00:05,80010C840,800122AA0,80010CF20,80012DA00,8834.09,Euro,8834.09,Euro,ACH,0

2019/01/01 00:08,80010CF20,80012DA00,80010CF20,80012DA00,9682.16,US Dollar,8262.75,Euro,ACH,0

2019/01/01 00:08,80010CF20,80012DA00,80010BD60,80011E460,9682.16,US Dollar,9682.16,US Dollar,ACH,0

2019/01/01 00:03,800319940,800466670,80029A010,8002F6F20,9125.22,US Dollar,9125.22,US Dollar,ACH,0

Data Info

Box: https://ibm.box.com/v/AML-Anti_Money_Laundering-Data

Github: https://github.com/IBM/AML-Data

- Data for one month: *January 2019*
- Size Compressed: *839,740,068 bytes*
- Size Uncompressed: *4,532,100,611 bytes*

- Number of Laundering Transactions = 55,423 *(0.12% of transactions)*
- Number of Non-Laundering Transactions = 45,351,509 *(99.88%)*

- Number of Bank Accounts = 3,935,305
- Number of Credit Card Accounts = 338,854

Related Work

- **Alternate AML Data and Models: https://github.com/IBM/AMLSim**
  - Focus on models
  - Data generation does not use detailed virtual world approach used here

- **Synthetic credit card transaction data: https://github.com/IBM/TabFormer**
  - Labeled for fraud / not-fraud
  - Github site contains both data and Transformer-based fraud detection model