bidities), and another organisation recording the survival data. Computing the log-rank statistic, however, requires knowledge of both the patient group information and the survival data. Lacking either type of data makes it impossible to deduce any meaningful insights. Using the cryptographic concepts of Secure Multi-Party Computation (MPC), it is possible to perform the analysis in this scenario while preserving the patients' privacy.

MPC is a set of techniques that enables multiple entities to jointly evaluate a function on their data, without revealing that data to one another. Some techniques achieve this property by supporting computations on encrypted data (e.g., homomorphic encryption), which particularly enables computations that involve the sensitive but encrypted data of another entity, whereas some other techniques (e.g., secret sharing) split the sensitive data in multiple pieces in such a way that computations can be performed on the separate pieces. Most importantly, within some specified security model, every MPC technique guarantees to preserve privacy throughout the entire computation.

In our joint 2020 research project, we have developed and implemented new MPC solutions to compute the log-rank statistic of the Kaplan-Meier estimator on vertically-distributed data. These privacy-preserving solutions do not reveal the group information and the survival data to anyone. Experiments show that the solutions are sufficiently fast and scalable to be used in real-world settings. The protocol is visualised in Figure 1. An open-source implementation is provided on GitHub [L2]. Our protocol does not reveal the Kaplan-Meier estimators themselves since patient-level information can be deduced from them. Presenting the Kaplan-Meier estimators in a more privacy-friendly way is described in Vogelsang et al. [1].

Motivated by these promising results, TNO and IKNL developed other relevant algorithms to enable privacy-preserving survival analyses, including Cox Proportional Hazard. Future activities include the extension of the toolkit to other relevant privacy-preserving machine learning algorithms for medical analyses in the cancer domain.

Links:
[L1] https://vantage6.ai/
[L2] https://kwz.me/h6S

References:

[1] Vogelsang et al.: "A Secure Multi-Party Computation Protocol for Time-To-Event Analyses", in Studies in health technology and informatics, 270, 2020, doi: 10.3233/SHTI200112.

Please contact:

Bart Kamphorst
TNO, the Netherlands
bart.kamphorst@tno.nl

Privacy-Preserving Collaborative Money Laundering Detection

by Marie Beth van Egmond, Thomas Rooijakkers and Alex Sangers (TNO)

Criminal transaction flows can be obfuscated by spreading transactions over multiple banks. Collaboration between banks is key to tackling this; however, data sharing between banks is often undesirable for privacy reasons or is restricted by legislation. In the MPC4AML project, research institute TNO and Dutch banks ABN AMRO and Rabobank are researching the feasibility of using Secure Multi-Party Computation (MPC) to detect money laundering.

Financial crime is a huge, world-wide problem. In 2018, an estimated 5.8 trillion dollars (6.7% of global GDP) worth of financial crime was perpetrated. Banks have a gatekeeper role in the financial system and a legal obligation to identify unusual transactions. However, criminal transaction flows will hardly ever stay confined to the network of one bank and thus often remain undetected. Therefore, collaboration and data sharing is key in detecting financial crime. On the other hand, legislation such as GDPR and competition law, as well as privacy concerns, can restrict data sharing.

Privacy-enhancing technologies are promising to enable collaboration without sharing sensitive data. This is also recognised by the recently published Future in Financial Intelligence Sharing (FFIS) paper [1] that describes ten case studies on this topic. One of these is the MPC4AML project, a collaboration between research institute TNO and Dutch banks ABN AMRO and Rabobank, where the technical feasibility of using Secure Multi-Party Computation (MPC) for detecting money-laundering is being researched. The possibility of performing calculations on the entire transaction network while keeping data private creates a lot of opportunities. In 2017, TNO published an article on secure PageRank for detecting transactional fraud [2]. In the current project, the research is focused on both the secure use of graph embeddings for finding malicious communities in the network, and on secure risk propagation for identifying exposure to high-risk cash or cryptocurrency deposits. We elaborate on the latter in this article.

Risk propagation

Currently, many banks attribute risk scores associated with money laundering to their customers, for example, based on transactions involving large amounts of cash or cryptocurrencies, or being from or to certain high-risk countries. These risk scores are of limited value as long as they are based on only the local network of a single bank. However, by propagating this risk through the transaction network of multiple banks, a lot more information on criminal flows could be identified. An example of such a criminal flow can be found in Figure 1.

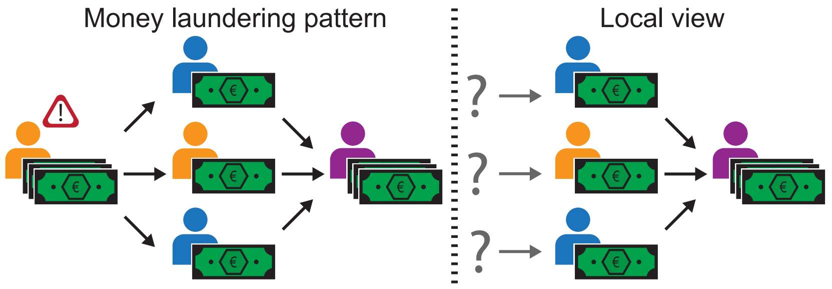

Figure 1: We consider a three-bank scenario (Orange, Blue, and Purple). In this scenario the first (left) account at bank Orange is classified as high risk (due to e.g., large cash deposits) by bank Orange. This account wishes to launder its resources. To stay under the radar, the resources are funnelled through multiple accounts, at various banks, before arriving at their eventual destination, e.g., the account at bank Purple (right). To detect money laundering, we wish to follow (propagate) the risky money and classify the endpoint as high risk too. Full (global) knowledge of the network enables us to propagate the risk. However, how can we achieve something similar when there is only partial (local) knowledge of the entire network available? This is where MPC comes into play.

$$r_k^j = (1-\delta) r_{k-1}^j + \frac{\delta}{T_j} \cdot \sum_{i \in S(j)} r_{k-1}^i \cdot A_{i,j}$$

○ r_k^j is the risk score of node j at iteration k,

○ δ is a public parameter between 0 and 1,

○ $S(j)$ is set of nodes linking to node j ,

○ $A_{i,j}$ is the transaction amount that node i sends to node j,

○ T_j is the total amount that node j receives.

Figure 2: Formula for risk propagation. Because of the properties of Additive Homomorphic Encryption, this can be performed on homomorphically encrypted risk scores as well.

The goal of risk propagation is to identify nodes that are involved in such a money-laundering pattern. The risk propagation algorithm requires as input a transaction network (containing clients and transactions) of a certain time period. Every client (a node in the network) has an initial risk score, which was assigned by the client’s bank. In one iteration of the algorithm, risk scores are updated using the weighted incoming risk score, i.e. the risk scores of incoming nodes weighted by the transaction amounts. In other words, if a client receives a lot of money from a client with a higher (or lower) risk score, its risk score will increase (or decrease). The formula for risk propagation can be found in Figure 2.

Secure risk propagation

Criminals often obfuscate money laundering patterns, such as in Figure 1, by performing transactions out of sight of any single bank. However, the risk scores that we need for the risk propagation are sensitive information that cannot freely be shared with other banks. Fortunately, using MPC techniques, the risk propagation algorithm can be performed securely on the entire transaction network of the participating banks. Risk scores are encrypted using Additive Homomorphic Encryption (AHE); a form of encryption that enables computations on encrypted data. To be precise, if we denote $[x]$ to be the encryption of a value x, additive homomorphic encryption has the property $[a] \bullet [b] = [a + b]$ and therefore $[a]^c = [c \bullet a]$.

To perform a secure risk propagation iteration, banks that participate in the protocol need to share the (relevant) encrypted risk scores with each other. Thanks to the properties of AHE, all banks can perform the required computations for risk propagation (in the formula in Figure 2) on these encrypted risk scores. Next to that, every bank can also use locally known information for each of its own clients, which enables a very efficient protocol. The result of the secure computation is an encrypted updated risk score for every client. With these updated scores, either a new iteration can take place, or a bank can decrypt the resulting risk score of some of its own clients. The latter can only happen with consent of the other banks, using “threshold decryption”.

Results

In collaboration with ABN AMRO and Rabobank, TNO built a first proof-of-concept of the secure risk propagation [L1] using synthetic data [L2] that includes money laundering patterns. This shows that it is possible to securely perform the risk propagation algorithm among different banks and demonstrates the value of collaborative transaction analysis.

Future challenges

Two technical challenges have been identified for the near future. First, the computational scalability of the solution will need to be evaluated and compared with the theoretical hypothesis. Second, MPC protects the confidentiality of the input data and any intermediate data during the computation. However, for application on real transaction data, it is important to investigate how much information can be deduced from the resulting risk scores after running the protocol.

The next step is to apply the secure risk propagation on real transaction data in a pilot. Many other challenges will no doubt arise from a pilot on real data. This will bring us one step closer to catching criminals involved in sophisticated interbank money laundering patterns.

Links:
[L1] https://kwz.me/h6q
[L2] https://github.com/IBM/AMLSim

References:

[1] N. Maxwell: “Innovation and discussion paper: Case studies of the use of privacy preserving analysis to tackle financial crime”, Future of Financial Intelligence Sharing (FFIS) research programme, 2020.

[2] A. Sangers, et al.: “Secure multiparty PageRank algorithm for collaborative fraud detection", Int. Conf. on Financial Cryptography and Data Security, Springer, 2019.

Please contact:
Marie Beth van Egmond,
TNO, The Netherlands
marie_beth.vanegmond@tno.nl

Alex Sangers
TNO, The Netherlands
alex.sangers@tno.nl