Securing the SWIFT Ecosystem

Exploring the Potential of Synthetic Data Applications

Whitepaper by

Pranav Satheesan, Lead Data Scientist - Mphasis NEXT Labs | Solution development
Mayank Umrao, Lead Data Scientist - Mphasis NEXT Labs | Solution development
Rahul Gupta, Manager (Data Science), Mphasis NEXT Labs | Solution lead
Uday Tangs, Fellow, Mphasis | Advisor
Hareesh PS, AVP, Mphasis NEXT Labs | Advisor
Biju Mathews, Partner - Mphasi s| Advisor

Mphasis
The Next Applied

Contents

1. Introduction: Why Synthetic SWIFT messages?

SWIFT (Society for Worldwide Interbank Financial Telecommunications) messages, a cornerstone of global financial communications, facilitate rapid, secure, and accurate cross-border payments. SWIFT does not process payments; instead, it provides a reliable messaging infrastructure to facilitate financial institutions worldwide to exchange standardized financial messages.

For instance, when a sender with an account in Bank A(say) in the USA wants to transfer funds to a recipient in Italy with an account in Bank B(say), the sender of funds must provide complete details about bank account information in Bank A, along with the recipient's bank account as well as the SWIFT code details of Bank B. After receiving the pre-transaction details along with the SWIFT Code of the recipient bank, Bank A sends a payment transfer message to Bank B through the SWIFT network, and upon receipt of this SWIFT message, Bank B processes the transaction.

Three factors drive the implementation of synthetic SWIFT messages in financial services. First, in the context of regulatory compliance testing, financial institutions employ synthetic data to mirror the characteristics of actual data without sensitive information to ensure continuity in systems and processes while maintaining customer privacy. Second, synthetic SWIFT messages are beneficial for training purposes, enabling staff to learn about processing financial transactions without the risks associated with using real data. Third, synthetic messages facilitate scenario testing, statistical and machine learning modeling, allowing institutions to simulate high transaction volumes or unusual transaction patterns for fraud detection without impacting real accounts.

This paper focuses on synthetic messages for the MT103 message type. Financial institutions send MT103 type of SWIFT message on behalf of the ordering customer, either directly or via correspondent(s), to the financial institution of the beneficiary customer. MT103 message type is utilized for clean payment instructions where the ordering customer, beneficiary customer, or both are non-financial institutions from the sender's perspective.

SWIFT standardized a format to encapsulate a transaction using structured information with identifiers in a SWIFT message to ensure the correct routing and processing of the transaction, provide necessary details for regulatory compliance, and enable the reconciliation of transactions. The message encompasses various fields, such as message type, application identifier, service identifier, and transaction reference, each providing different pieces of information about the transaction. Some common identifiers within a SWIFT message include sender and receiver information, transaction details, banking institution details, beneficiary information, payment instructions, charges and fees, regulatory and reporting information, and remittance information.

2. The Need for Synthetic SWIFT Messages

While the SWIFT system and synthetic SWIFT messages offer significant advantages to the financial industry, the following five business challenges associated act as a motivation to create a robust system to handle synthetic SWIFT messages.

- Ensuring Compliance: Financial institutions need to ensure that their systems comply with the regulations and standards of SWIFT. The complexity and diversity of these requirements make it challenging to develop and maintain compliant systems.
- Data Privacy: Protecting the privacy of customer data while conducting tests or downstream activities is a critical concern. The challenge lies in generating synthetic data that closely mimics the original data but does not contain any sensitive information.
- Training and Education: It is necessary to provide effective training on the handling and processing of SWIFT messages. However, using real data for training purposes can pose risks, making it challenging to create realistic training scenarios without compromising data security.
- Scenario Testing: Simulating various scenarios, such as high transaction volumes, unusual transaction patterns for fraud detection, or specific error conditions, is crucial. The challenge lies in the creation of synthetic messages that accurately represent these scenarios without affecting real accounts.

- Decoding SWIFT Messages: The structured format of SWIFT messages, which includes various identifiers and fields, can be complex to decode and understand. This poses a challenge for the staff dealing with these messages, especially in the case of manual processing or error resolution.

3. The Context and the Problem Statement

The principal technical challenge is generating synthetic SWIFT messages with the same structure and information as the original message while ensuring data privacy and security. The task is complex due to two primary reasons.

First, the original dataset contains mandatory fields, with optional tags needing to be created based on the rules defined in the SWIFT technical documentation.

Second, a significant amount of Personally Identifiable Information (PII) must be masked while generating synthetic data to ensure the privacy of individuals and institutions.

Both challenges are difficult to solve because the process of synthetic SWIFT message data generation involves understanding the complex structure of SWIFT messages, implementing the rules to generate optional tags, and masking the PII effectively without affecting the usability of the data. Moreover, the solution needs to be robust enough to handle different types of SWIFT messages and flexible enough to adapt to changes in the SWIFT message format or rules.

The traditional approach to analyze SWIFT messages is to take the database of actual SWIFT messages, mask out any PII, and then feed it into the model. Although this method ensures data privacy, it is tedious, error-prone, and requires a significant amount of human effort. Moreover, it may not fully address the problem of creating a realistic synthetic dataset that can be used for testing, development, training, and scenario testing in the financial industry.

To overcome the challenges, a novel approach is developed to generate synthetic SWIFT messages of MT103 message type. This approach creates tags synthetically, reduces human effort, increases data privacy and security, and provides a realistic dataset for various uses in the financial industry. The proposed solution can be customized for other types of SWIFT messages.

4. Our Solution

Mphasis integrated a novel algorithm in Mphasis Synth Studio to generate synthetic SWIFT messages as a part of an Enterprise Synthetic Data Platform to cater to the premium synthetic data requirements of enterprises. The data from Mphasis Synth Studio serves as a potent source for testing several internal and external applications and at the same time, preserving privacy and safeguarding data subjects.

The uniqueness of the solution lies in its capability to build accurate and reliable models, even in situations where the availability of data is low or entirely non-existent. Synth Studio offers Synthetic data for SWIFT MT103 (Single Customer Credit Transfer), a data product that further enriches its wide array of offerings.

The SWIFT MT103 message comprises several tags, both mandatory and optional, that contribute to its structure. The tags are mentioned below:

- Tag 23: Bank Operation Code: specifies the transaction type.
- Tag 32A: Value Date, Currency, Amount: indicates the date when the funds become available, the transaction's currency, and the amount.
- Tag 33B: Currency/Original Ordered Amount: specifies the currency and the original order's amount.
- Tag 50: Ordering Customer: identifies the party initiating the transfer.
- Tag 52(A/D): Ordering Institution: represents the bank from where the payment has been sent.
- Tag 53(A/B/C): Sender's Correspondent: the bank that is acting on behalf of the sending institution.
- Tag 54(A/B/C): Receiver's Correspondent: the correspondent bank for the receiver's side.
- Tag 56(A/C/D): Intermediary Institution: a third-party bank that facilitates the transfer.
- Tag 57(A/B/C/D): Account with Institution: the bank where the beneficiary's account is held.
- Tag 59: Beneficiary Customer: the party that is receiving the funds.
- Tag 70: Remittance Information: details about the payment's purpose.

5. Methodology: Preprocessing and Modelling Approaches

The initial step in the solution involves creating data for optional fields, adhering to the rules of MT103 and capturing all possible combinations by creating a database comprising the top 10 banks from the top 20 economies. Both mandatory and optional tags are formulated to mimic message transfer between banks in these countries.

The Conditional Tabular Generative Adversarial Networks (CTGAN) model was chosen as it can generate synthetic data by learning the underlying distribution of features in the dataset.

Preprocessing of the data ensures consistency across various sub-tags corresponding to the same tag which is done by enforcing a Fixed Combinations constraint on sub-tags of the same tag which ensures that the model maintains consistency at the sub-tag level while generating synthetic tags.

The performance of the synthetic data is assessed based on the following metrics:

- Column Shapes: This metric measures the statistical similarity between the real and synthetic data for single columns of data.
- Column Pair Trends: This metric measures the statistical similarity between the real and synthetic data for pairs of columns.

These metrics ensure that the model not only learns the distributions of features (i.e., tags/sub-tags) but also the interaction between tags. Consequently, the model can learn both intra tag/sub-tag and inter tag/sub-tag distributions.

As a note of caution, users intending to use the model with filtering countries/banks or using custom data should strive to maintain the value of these parameters above 50%. A good generalizable model typically exhibits values between 70-90% on both these metrics.

6. Data Validation

Synth Studio has custom functions for validating synthetically generated data. We calculate ‘par score’ to check if the synthetic data doesn’t contain original data points. First, pairwise Euclidean distance between a pair of points from real and synthetic datasets are taken which is termed as ‘external distance’. If the distances are relatively small, it suggests that the synthetic data closely resembles the original data. Larger distances indicate greater dissimilarity. Then, pairwise Euclidean distance between a pair of points from the real dataset is also taken, which is termed ‘internal distance’. The ratio of internal distance to the external distance is named ‘lift’. Ideally the lift should be always greater than 1 for each datapoint. The points having lift greater than 1 are termed ‘close’ points. The ‘par score’ is the mean lift of all ‘close’ points in the data. It is a hyperparameter and the acceptable ‘par score’ is empirical. In our experimentations, we found that a generalizable model typically exhibits values between 70-90% par score. Users can manually examine points which have a lift greater than or equal to 1 to check for any instance of real data in the generated data.

The solution provides users with the flexibility to customize the solution according to their requirements. They can filter out countries and/or banks to mimic transactions among specific countries/banks or geographies. Additionally, users can input custom banks for synthetic data generation, which can help generate synthetic data specifically curated for their use case.

The solution gives developers creating products in the SWIFT ecosystem four major superpowers:

- Access to high-quality synthetic data which mimic actual SWIFT messages
- Reduced developer time for creating diverse solutions
- Reduced time and effort for data exploration and ingestion
- Reduced cognitive load on verifying possible exposure of PII information

7. The Impact of the Solution

The proposed solution has the potential to significantly enhance customer experience in a multitude of ways.

The generation of high-quality synthetic SWIFT MT103 messages helps financial institutions to conduct rigorous tests and simulations without compromising the privacy of their clients. This ensures that the systems are foolproof and reliable, leading to fewer errors, faster processing times, and improved customer satisfaction.

The solution enables developers working in the SWIFT ecosystem access to premium synthetic data that mirrors actual SWIFT messages and makes the development process more efficient and accurate. They can test their products in realistic scenarios without the need to use real customer data. The solution also reduces their cognitive load as they no longer need to worry about possible exposure of Personally Identifiable Information (PII), resulting in a safer and more secure product development environment.

The solution also offers customization, allowing users to tailor the data generation process to their specific needs. This can be particularly useful for institutions looking to understand transaction patterns between specific countries, currencies, or banks. Such a level of customization can be invaluable in enhancing the institution's services and offerings. The synthetic data generated can also be used for training and education purposes by training new employees about the handling and processing of financial transactions without the risk associated with using real data. This will ensure that they are better prepared to handle real-world scenarios, leading to improved efficiency.

The solution's ability to create synthetic SWIFT messages has significant implications for improving customer experience. By ensuring secure, reliable, and efficient transactions, it contributes to the building of trust and confidence in the financial institutions leveraging this technology. It offers a win-win situation for all stakeholders involved - the financial institutions, their clients, and the developers working to enhance the SWIFT ecosystem.

8. Scope of Future Work

The solution could be extended to other message types using Synth Studio. The steps involved will remain same: creation of data using rules and then training the model to learn inherent patterns in the data, followed by data validation. The solution must be customized to other message types, basis the rules involved in creating the tags for that specific message type.

For instance, message type MT-202 is a message carried for General Financial Institution Transfer. It is used in cases when all parties involved in a transfer are financial institutions only. There are tags which have the same name, but different contexts across both message types:

Tag 52: Ordering Institution

- MT-103: Represents the bank from where the payment has been sent on behalf of the customer.
- MT-202: Represents the bank initiating the interbank transfer.

Tag 53: Sender's Correspondent

- MT-103: Specifies the bank that is acting on behalf of the sender.
- MT-202: Specifies the intermediary bank settling funds for the sender bank.

Tag 54: Receiver's Correspondent

- MT-103: Denotes the correspondent bank for the receiver's side.
- MT-202: Denotes the bank handling settlement for the recipient bank.

Tag 56: Intermediary Institution

- MT-103: Identifies a third-party bank that facilitates the transfer for the sender.
- MT-202: Identifies the intermediary bank in the chain of interbank settlements.

Tag 57: Account with Institution

- MT-103: Specifies the bank where the beneficiary's account is held.
- MT-202: Specifies the final receiving bank in the interbank transfer process.

These tags have the same format, hence generation of synthetic data for these tags would follow same principle. The data would look different as the same tag in MT-103 would mimic a single customer's identifier properties while for message type MT-202, it would mimic a financial institution's properties. These nuances must be taken care of while generating synthetic data.

9. Conclusion

The application of synthetic data, specifically synthetic SWIFT messages, has profound implications for the financial industry.

It addresses the challenges associated with data privacy, training, compliance, and scenario testing, all while maintaining the integrity and functionality of the original data. This solution, developed by Mphasis Synth Studio, provides a robust system for creating synthetic SWIFT messages, specifically the SWIFT MT103 messages.

The solution allows for rigorous testing, simulations, and training without compromising the privacy of clients, ensuring the creation of reliable, efficient, and secure financial systems. It not only enhances the experience of the customers but also benefits developers by providing them with high-quality synthetic data, reducing development time and mitigating the risk of exposure of PII information. Moreover, the solution offers customization to cater to the specific requirements of users, providing deeper insights into transaction patterns and contributing to the improvement of services and offerings.

In essence, the ability to create synthetic SWIFT messages not only ensures secure, reliable, and efficient financial transactions but also builds trust and confidence among all stakeholders - the financial institutions, their clients, and the developers working in the SWIFT ecosystem. The advent of synthetic data signifies a paradigm shift in how data is perceived and utilized, paving the way for innovative solutions that are compliant, secure, and customer-centric.

The adoption and implementation of synthetic data are indeed the way forward in the rapidly evolving digital age. The solution can be accessed from the links provided below.

https://aws.amazon.com/marketplace/pp/prodview-nsfhelo72jwrw

About Mphasis

Mphasis' purpose is to be the *"Driver in the Driverless Car"* for Global Enterprises by applying next-generation design, architecture and engineering services, to deliver scalable and sustainable software and technology solutions. Customer centricity is foundational to Mphasis, and is reflected in the Mphasis' Front2Back™ Transformation approach. Front2Back™ uses the exponential power of cloud and cognitive to provide hyper-personalized ($C = X2C^{2}_{TM} = 1$) digital experience to clients and their end customers. Mphasis' Service Transformation approach helps 'shrink the core' through the application of digital technologies across legacy environments within an enterprise, enabling businesses to stay ahead in a changing world. Mphasis' core reference architectures and tools, speed and innovation with domain expertise and specialization, combined with an integrated sustainability and purpose-led approach across its operations and solutions are key to building strong relationships with marquee clients. Click here to know more. (BSE: 526299; NSE: MPHASIS)

For more information, contact: marketinginfo.m@mphasis.com

USA
Mphasis Corporation
41 Madison Avenue
35th Floor, New York
New York 10010, USA
Tel: +1 (212) 686 6655

UK
Mphasis UK Limited
1 Ropemaker Street, London
EC2Y 9HT, United Kingdom
T : +44 020 7153 1327

INDIA
Mphasis Limited
Bagmane World Technology Center
Marathahalli Ring Road
Doddanakundhi Village, Mahadevapura
Bangalore 560 048, India
Tel.: +91 80 3352 5000

WAS 10/07/24.US.LETTER.BASIL.9152

www.mphasis.com