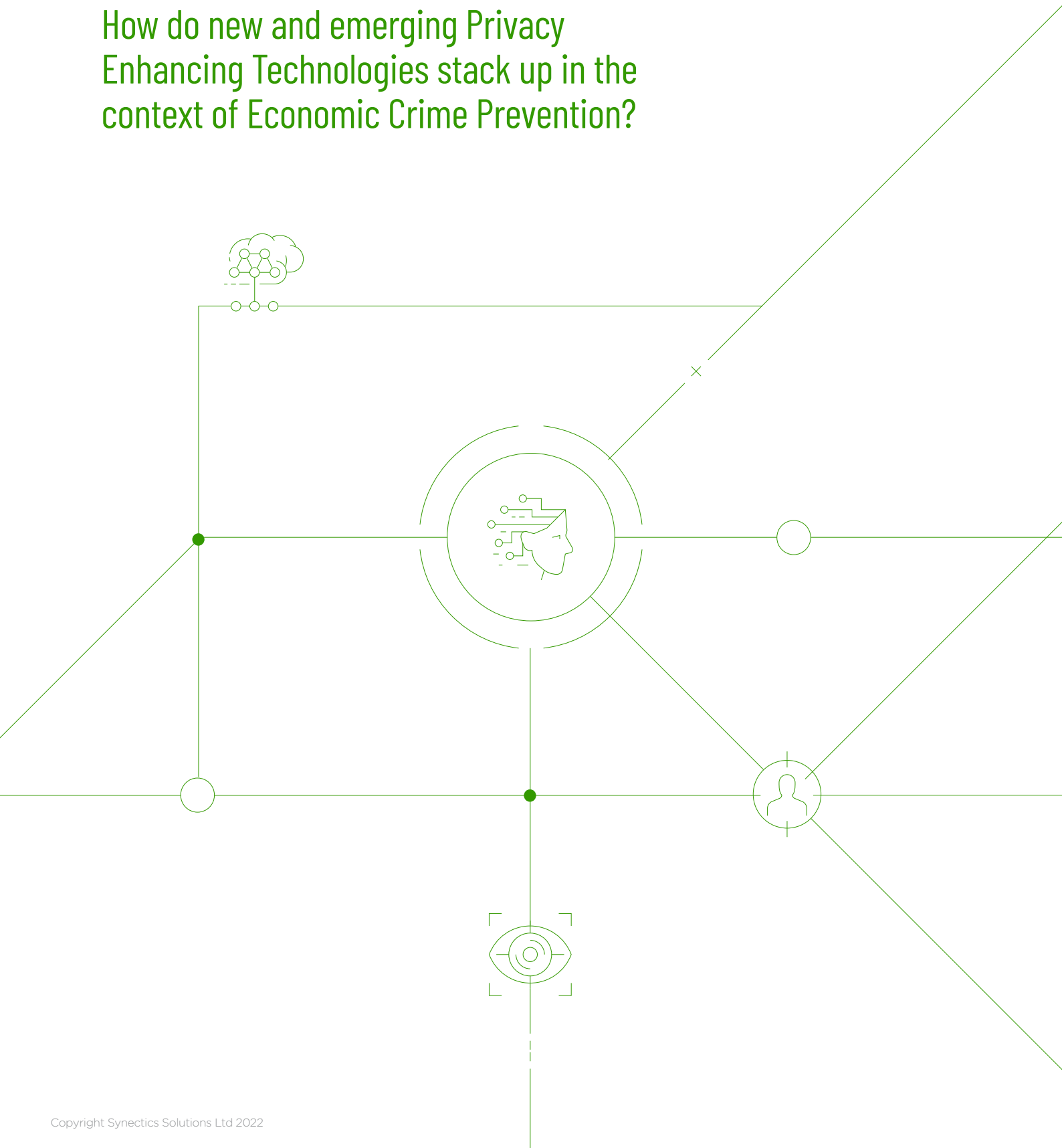
SYNECTICS SOLUTIONS

The Secret Life of PETs

How do new and emerging Privacy Enhancing Technologies stack up in the context of Economic Crime Prevention?



Foreword

This paper is an important, balanced and timely contribution to the ongoing debate about the place of privacy enhancing technologies in the fight against fraud and financial crime.

There is a significant trade off when considering introducing such technology, with enhanced privacy needing to be weighed against reduced efficiency and increased cost and computational capacity. However, it is clear that in time and in particular use cases privacy enhancing technology will take its place alongside other existing effective fraud prevention technologies.

Mike Haley
CEO, Cifas

In the UK, please contact:

Synectics Solutions
Synectics House
The Brampton
Newcastle-under-Lyme
Staffordshire
ST5 0QY

Tel: (01782) 664000

Synectics Solutions website:
www.synectics-solutions.com

The Secret Life of PETs – How do new and emerging Privacy Enhancing Technologies stack up in the context of Economic Crime Prevention?

A REVIEW OF THE CURRENT LANDSCAPE FOR PETS

OPPORTUNITIES THAT PETS PRESENT IN THE CONTEXT OF ECONOMIC CRIME PREVENTION

AN EVALUATION OF COMMON EXAMPLES OF PETS AND HOW THEY STACK UP IN PRACTICE

CHALLENGES AND LIMITATIONS

CONCLUSION

New and emerging Privacy Enhancing Technologies (PETs) represent a significant opportunity for data analytics – promising the ability to perform complex calculations, aggregations and queries on disparate data sets, without compromising the security of personal data in any fashion.

In Gartner Hype Cycle terms, new PETs have passed the "Peak of Inflated Expectations" and are, like blockchain, now in the pre-production stage for various prototypes and proofs of concept for different use cases.

A particular area ripe for exploration is in the world of fraud prevention and financial crime risk analysis - a significant market which carries both major investment and major challenges regarding data security, financial regulation and competitive risk.

However, as with any new technology, there are potential pitfalls that can be overlooked in favour of technological panacea.

Compatibility with legacy systems, access to skilled technical and analytical resource, and infrastructure cost to name just a few!

Notwithstanding the complex regulatory, legislative and operational environments that currently make up the global financial services landscape.

In parallel, the Financial Action Task Force (FATF) have stated that sharing data to prevent fraud and financial crime[1] (AKA Economic Crime)[2] is essential within organisations, across organisations and across territories.

The wider geopolitical landscape necessitates an acceleration of this process. There is a key requirement to share more data, more frequently and at pace to prevent, detect and disrupt organised crime. All whilst maintaining privacy and data security for regular citizens – a difficult balancing act that PETs may help address.

The facts, however, remain stark – multiple £trillions of illicit funds flow through global financial services annually according to various sources, including the United Nations Office on Drugs and Crime[3].

But the monies recovered from serious and organised crime, actions of nation states and even opportunistic criminals are far outweighed by cost.

This paper will evaluate a selection of PETs in the context of fraud and financial crime risk analysis, to help determine what opportunities PETs might provide, but also where traditional analytical and data matching techniques could be favourable as the PET landscape matures.

1 – https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-information-sharing.html
2 – https://www.gov.uk/government/publications/economic-crime-plan-2019-to-2022
3 – https://www.icij.org/investigations/fincen-files/global-banks-defy-u-s-crackdowns-by-serving-oligarchs-criminals-and-terrorists/#:~:text=Estimates%20by%20the%20United%20Nations,of%20the%20world's%20dirty%20money.

The Current Landscape

The Future for Financial Intelligence Sharing think tank, in collaboration with RUSI, has completed an evaluation of the current private sector financial intelligence sharing landscape entitled **Lessons in private-private financial information sharing to detect and disrupt economic crime.**[4]

This document describes current systems that enable fraud and financial crime-related data sharing.

Using data sharing for good

Some data sharing platforms, such as our own National SIRA platform[5] and the CIFAS National Fraud Databases[6] (both UK-based), are well established and respected with hundreds of contributing organisations.

They provide point-in-time risk data to support fraud investigations, network analytics and wider risk management for the financial services and insurance sectors.

In contrast, newer initiatives, such as the Transaction Monitoring Netherlands POC, are more nebulous and tactical in their scope and maturity but may adopt some privacy enhancing mechanisms in their architecture.

The maturity of the UK counter fraud data sharing market – which benefits from clear regulatory, legislative and operational gateways that enable data sharing for the common good of preventing and disrupting criminality – is an exception rather than the norm.

UK Finance estimates that in UK financial services £2 out of every £3 of attempted fraud is prevented[7], whilst we (Synectics Solutions) save National SIRA members over £1billion annually in prevented fraud losses[8]. The UK economic crime intelligence ecosystem is a fundamental driver behind the City of London being a world leading cog in the global financial services landscape.

These existing platforms have core tenets of shared definitions between participating agencies and the principle of reciprocity – in that any member of a data sharing consortium must contribute information back proportional to the intelligence gained from other members.

Despite the strong case study represented by the UK approach, the aforementioned FFIS evaluation states "detection and investigation of economic crime has been stymied by analytical efforts being siloed and fragmented on many levels, including:

- **At the level of individual private sector institutions**
- **By business sector**
- **Between public and private sectors**
- **Between domains of economic crime**
- **Across borders"**

This is despite economic crime data sharing evidencing a number of advantages, summarised by FFIS as:

- **Analysis over broad data rather than silos**
- **Observation and assessment of risk**
- **Reducing duplicated efforts across institutions**
- **Enabling early prevention & disruption**
- **Enabling digital customer experience**

4 – Video summary here - https://www.youtube.com/watch?v=shYlqaVYUFg – full report to be published
5 – https://www.synectics-solutions.com/
6 – https://www.cifas.org.uk/
7 – https://www.ukfinance.org.uk/policy-and-guidance/reports-publications/fraud-facts-2021
8 – https://www.synectics-solutions.com/about-us

Red tape and wider economic factors

Another consideration in the current ecosystem are the various restrictions that financial institutions and the public sector have to operate within. This environment can limit their agility and ability to develop innovative new technologies to disrupt current processes.

Legacy systems, regulatory provisions, data governance considerations and limited resource are additional weights that can hold back the tide of innovation.

This is further exacerbated by the uncertain times we are currently in, including our emergence from the Covid-19 pandemic, the war in Ukraine, impact of global climate change and UK domestic issues such as Brexit.

Mature economic crime intelligence platforms work around these restrictions. Leading platforms use technology such as machine learning with clear regulatory and legislative gateways that encourage collaboration and data sharing, whilst protecting individuals' privacy through encryption, secure hosting environments and clear usage guidelines for data.

The restricted ability to integrate and innovate in an increasingly complex landscape will not only affect the adoption of Privacy Enhancing Technology –

see the glacial adoption of blockchain technology as another example of how theory does not always easily translate into reality.

Bringing cybercrime into focus

Alongside the economic crime landscape, the privacy landscape has also been changing. With the volume of data breaches continuing to grow year on year, cybercrime is a core focus for regulators, policymakers and financial institutions alike. Consumer attitudes to data privacy are hardening, and the proliferation of disinformation and misinformation is exacerbated by readily available compromised personal data.

In the UK, the Online Safety Bill[9] directly tasks social media websites with policing the content on their platforms or face the threat of fines. The need for secure data analytics continues to increase – not just to mitigate data breaches, but also to widen the ability to cut through noise and allow institutions to make data driven decisions without being over-encumbered by privacy obligations.

9 – https://www.gov.uk/government/news/major-law-changes-to-protect-people-from-scam-adverts-online

So – what are PETs and how could they improve the existing intelligence sharing ecosystem?

Secure Multi-Party Computation

According to AIMultiple.com, "secure multi-party computation (also called multi-party computation, SMPC, or MPC) is a cryptographic technique that enables different parties to carry out a computation using their private data without revealing their private data to each other.

A popular example to illustrate the basic idea behind SMPC is as the following:

Suppose a group of employees wants to learn their average salary in order to find out whether they are underpaid or not. However, they don't want to disclose their individual salary information. An SMPC method can solve this problem:

1. **Each employee is numbered from first to last.**
2. **The first employee chooses an arbitrarily large number and adds their salary to the number and tells the second employee the result.**
3. **The second employee adds their number to the value and tells the result to the third employee, and so on until the last employee.**
4. **After adding their salary to the result, the last employee tells the result to the first employee.**
5. **The first employee subtracts the large number they started with and divides the result by the number of employees in the group to obtain the average salary.**

In this example, the large number chosen by the first employee hides his/her salary from others. On the other hand, the final result that the first employee receives from the last employee provides no information to the first employee about others' salaries. As a result, the group, consisting of multiple parties, could securely compute the average salary without disclosing their salaries."[10]

SMPC: Explained

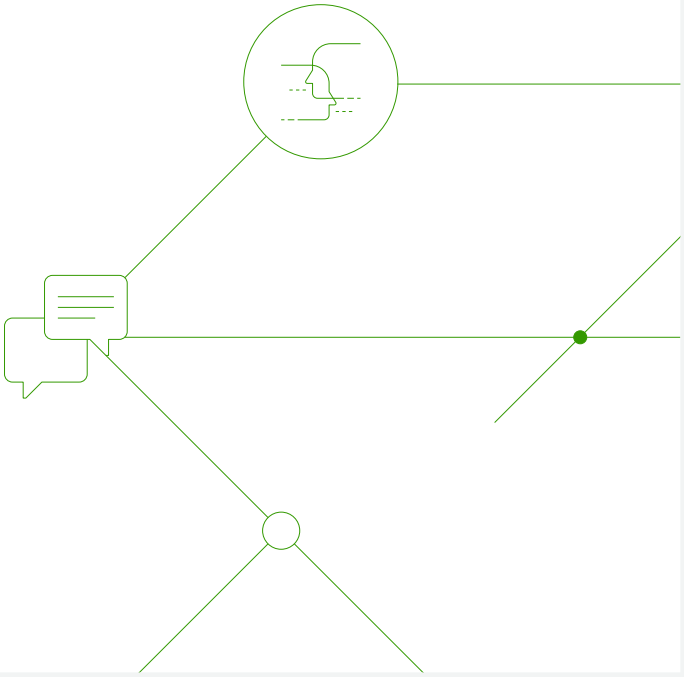
SMPC is designed to ensure each party participating in a computation only receives the necessary results pertinent to their business question. Underlying data, or "workings out", are segregated from any output.

Generally speaking, SMPC allows for distributed computation across multiple separately owned data sets without the need for a single party to see another's data, provided data is in a common format and has been sufficiently prepared to facilitate analytics.

Rather than a centralised data store, data is distributed over the various institutions (parties). Complexity in data preparation increases per institutions included in the computation, which can limit utility whilst increasing the cost of implementation and maintenance of the consortium.

When implemented correctly SMPC allows the parties to communicate insights without exposing underlying data. Parties only return results relevant to the data or questions they have contributed to the computation.

An example benefit would be to enable a Machine Learning model to be trained on multiple disparate private data sources without pre-aggregation of the data into a central store.



10 – https://research.aimultiple.com/secure-multi-party-computation/

Protecting yourself from potential pitfalls

Poor data in = poor insight out

There is a risk of malicious parties contributing incorrect or invalid information into the computation, as all coordinating parties are masked from each other within an SMPC consortium.

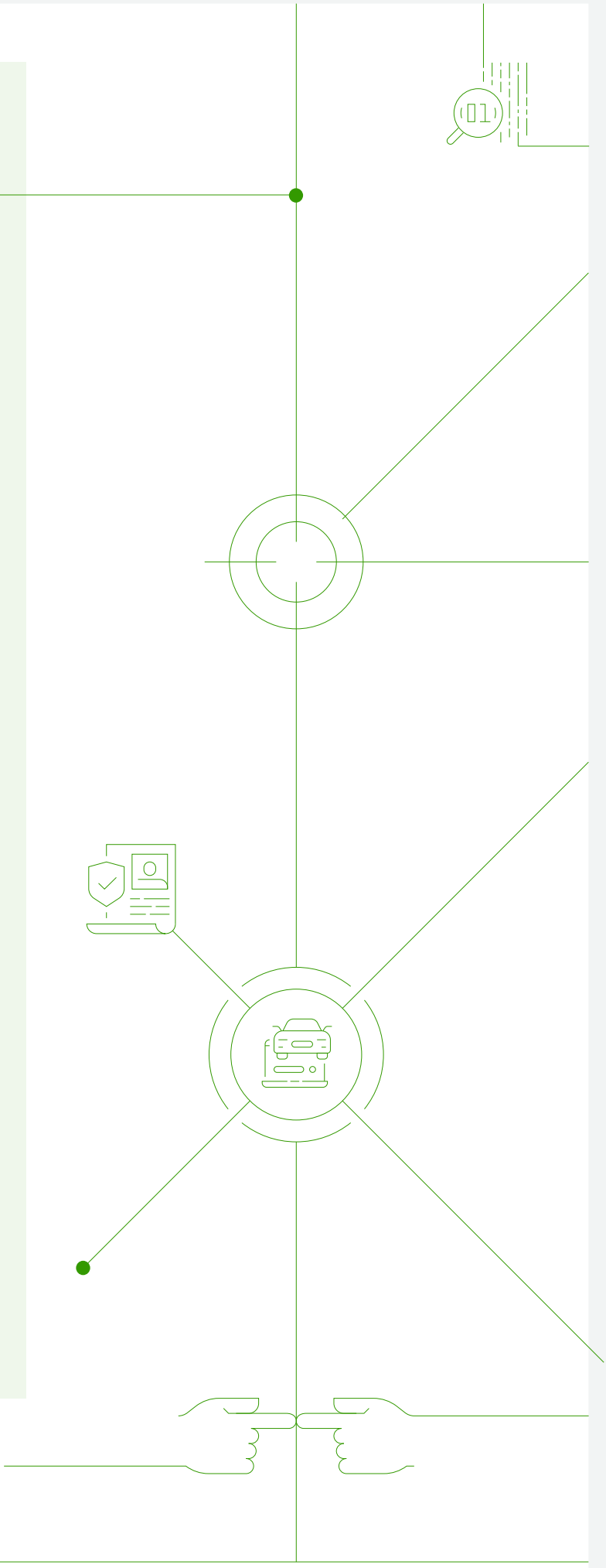
In the example above, all it takes is for one employee to lie about their salary and the solution is void. This is unavoidable due to the nature of SMPC and adds regulatory considerations to any practical implementation of the technology.

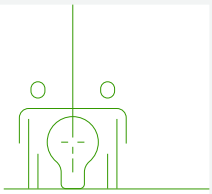
SMPC carries a higher computational cost (one or more "orders of magnitude" according to one leading company in the sector) than traditional analytical methods, and requires all parties to contribute to the computation. In practice, this may mean some parties are disproportionately affected dependent on the volume of requests versus the size and quality of the data they have contributed.

SMPC in its current state appears to be best placed for tactical/specific incidences of consortium analytics between a finite list of trusted parties. Questions should be predefined dependent on data contributed to the computation and any data preparation that has taken place upstream.

In terms of the existing infrastructure for sharing economic crime intelligence, the clearest use case is for cross-border analytics across multiple distinct intelligence databases, particularly to support machine learning use cases.

As fraud prevention is continually evolving and benefits from increased complexity in underlying data, alongside increased volume in reference sources, SMPC becomes cost prohibitive and unwieldly as an alternative to traditional data matching techniques.





Zero-Knowledge Proof

Zero-Knowledge Proof (ZKP), a wider concept of which SMPC is an application thereof, is difficult to describe simply, evidenced by the popular "How to explain Zero-Knowledge Proof to your children" Ali Baba example being six pages long and incredibly complicated[11].

A one-line description would be: "a Prover wants to convince a Verifier that a statement is true without revealing any further information."

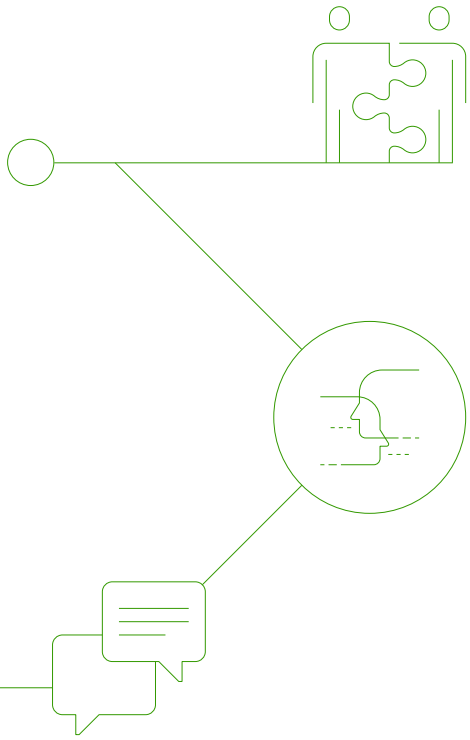
ZKP: explained

I have found that the most effective way to describe Zero-Knowledge Proof in non-mathematical terms is the "two balls and a colour-blind friend description", which I have summarised below:

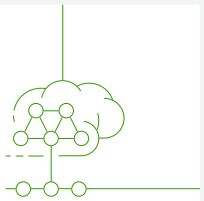
- My friend (let's call him Russell) is red-green colour blind. I have a red ball and a green ball, which aside from the colours, are otherwise identical. These balls look identical to Russell.
- I wish to convince Russell that the balls are different colours.
- I give the balls to Russell who puts them behind his back. He pulls one ball from behind his back and holds it in front of him, so I can see it. He puts it behind his back. He then does this again and asks me whether he has switched the ball.
- This procedure is repeated as often as required.
- By looking at the colours, I can say with certainty whether Russell has switched the ball, and after enough times, the probability of having guessed the correct colour each time becomes inconceivably small.
- Russell now knows the balls are different colours (proof) but does not know which ball is which colour (zero knowledge).[12]

Zero-Knowledge Proof can be seen as an extension to Secure Multi-Party Computation, as SMPC needs to be in place in order to ask ZKP questions.

A practical use case is to ask a number of financial institutions as to whether a new customer to a particular institution has been listed as a PEP at any other institution, without disclosing the customer's personal details or which other institutions the customer banks with.



11 – Accessed here for the curious: https://pages.cs.wisc.edu/~mkowalcz/628.pdf
12 – A more detailed description: https://www.linkedin.com/pulse/demonstrate-how-zero-knowledge-proofs-work-without-using-chalkias/



Homomorphic Encryption

Homomorphic encryption is a type of data encryption designed to allow mathematical operations to be performed on encrypted data.

Homomorphic encryption: explained

Essentially, it enables two encrypted data points to be operated upon (such as multiplied) in a fashion that when decrypted, would have the same result as performing the same operation on the data points in plaintext.

This includes the ability to match data within a fraud data matching scenario or screening individuals against PEPs and Sanctions lists, should it be that homomorphic encryption has been applied to all relevant data.

It also needs to be impossible to reveal information regarded the encrypted data points by observing encrypted calculations. So, in summary, incredibly complicated maths to enable encrypted regular mathematics.

As data is encrypted any data that has homomorphic encryption applied could be outsourced to a third party without necessarily trusting said third party to secure the data.

Protecting yourself from potential pitfalls

The main issue with homomorphic encryption is efficiency. According to KeyFactor[13], fully homomorphic encryption can be up to a million times more computationally intensive than performing operations in plaintext. These algorithms are slow and have significant storage requirements, which translates to both poor response times and high cost, particularly when considering cloud storage and compute costs.

When considering the incredibly tight service level agreements associated with fraud data matching, identity verification and PEPs/ Sanctions screening, homomorphic encryption currently represents an unacceptable trade-off in terms of performance. Especially when compared to traditional data matching and analytical techniques in the world of economic crime in the context of customer decisioning.

Homomorphic encryption is also incredibly complex in terms of mathematics, and incompatible with most operational systems used across financial services and adjacent sectors today, which represents a significant barrier to entry.

As KeyFactor conclude, “while homomorphic encryption may not be a viable option today, it’s possible that could change in the future”.

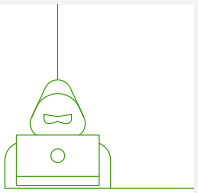
Other examples of Privacy Enhancing Technology

The three highlighted above aren’t the only examples of new or emerging PETs. Other technologies include:

- **Differential privacy - The process of adding noise, to ensure the output of statistical analysis on that data will not reveal information specific to a single individual from the dataset.**
- **Enclaved data - A secure data environment which limits access to confidential data at a hardware level, for example through the use of virtual machines, secure networks and protected memory regions. Data enclaves operate as opaque boxes to outside users and processes.**
- **Federated analytics - The execution of programs on decentralised data. Data remains in place and is not shared, with only the results returned to the requesting party.**
- **Synthetic data – Fake or computer-generated data designed to mimic real data, to train machine learning models, conduct mock analytical exercises or test production-systems without production data.**

As with any technology, there is overlap between the different technological approaches described, and if you are interested there is a wealth of information available about each approach freely available on the internet. Warning: the mathematics gets very hard, very quickly!

13 – https://www.keyfactor.com/blog/what-is-homomorphic-encryption/



Other examples of Privacy Enhancing Technology

PETs are not a silver bullet, and it is safe to say that no single PET will fully address privacy challenges present in today's data driven ecosystem. If we had the benefit of infinite resource and a clean slate, it may be different, but as with all new innovations this is not the case.

Challenges & Limitations of PETs

Broadly, new and emerging PETs should be applied within a wider "Privacy by Design" approach proportionate to both the relevant privacy risks associated with a particular activity and the anticipated utility and benefit of said activity.

Many of these challenges and limitations are the same as those associated with other innovations, such as blockchain technology, and prior innovations that are becoming widespread, such as artificial intelligence and machine learning.

This represents the state of play today and will evolve as technology matures and best practice is established for these new techniques. Synectics are actively pursuing R&D projects in this area to help inform debate.

Core challenges can be summarised as follows:

- **Technical expertise** – PETs are niche and require specific dedicated technical resource to support. As with premium quality data scientists, this is a limited pool of talent which can carry a high cost.
- **Financial cost** – PETs are currently expensive to run in operation, with many specialist vendors still in the start-up stage, not benefiting from the economies of scale found with more established analytical techniques and technologies.
- **Compute / infrastructure cost** – Particularly with techniques such as Homomorphic Encryption, increased storage requirements combined with increased demands on CPU processing power can massively inflate the cost of running PETs in production environments, particularly when considering multiple contributing parties and high data volumes.
- **Compatibility with legacy systems** – particularly in financial services, many core platforms are incompatible with new technologies and processes. This is not just an issue for PETs – compatibility issues also come into play for machine learning, blockchain and cloud computing.
- **Data preparation / data quality** – PETs still require high quality data to perform effectively. This, like analytics on clear text, normally requires extensive data preparation, particularly in a supervised machine learning-type environment. As data preparation is made more difficult by the application of a PET, this can mitigate the benefit of the PET in the first place.
- **Potential for misuse** – PETs could introduce transparency and accountability risks, and are often dependent on absolute trust in all participant parties. Bad actors could take advantage of this to use data in harmful or other unethical ways.
- **Fraud investigations –** complex fraud typologies, such as those prevalent in the Finance and Insurance sector, require tools such as Open Source Intelligence to support investigations and ultimately prosecutions. Certain analytical or comparison methods, for example checking similarities or differences between emails and personal details, also struggle when all PII is masked.

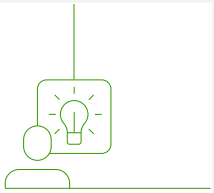
FORBES HAVE STATED THAT

84% of Digital Transformation projects FAIL...

due to the failed adoption of technology generically, and we all have examples within our own organisations where the introduction of a new technology or technologies has failed to deliver the promised results.

When this is overlaid with the threat of financial losses due to fraud, regulatory compliance obligations and poor customer experiences, the fear of change can become palpable. This is further compounded by a lack of widespread understanding of the complex technical concepts behind PETs, which currently suffers from a lack of clear referenceability in production environments.

Even Machine Learning, which has started to deliver on the initial promises of the prior decade, is still being adopted in the Economic Crime world at a glacial pace, and still feels far away from the standard for specific regulatory obligations such as PEPs & Sanctions screening for Customer Due Diligence.



All is not lost...

The privacy conscious reader might be feeling a bit deflated at this point, but there are reasons to be hopeful.

Existing established services for economic crime intelligence sharing already operate Privacy Enhancing Technology, such as tokenisation, various levels of encryption and secure user access controls. Using Synectics as an example, we are:

- **ISO27001 certified – the leading certification for information security and data governance**
- **Cyber Essentials certified – the leading government programme for cyber security**
- **A "Specified Anti-Fraud Organisation," named directly by the UK Government, with specific legislative provisions to provide public / private data sharing for the purposes of fighting fraud in the context of economic crime[14].**

As per GDPR, the Data Protection Act and guidance from the ICO, we:

- **Operate a "Privacy by Design & Default" policy for data, software and analytics**
- **We hold financial inclusion, social justice and data protection as core tenets to our business**
- **Only access, analyse and share data with clear contractual, security and governance provisions**
- **Gate access to data behind secure APIs**
- **Implement access controls for any system, including those that do not contain personal data**
- **Encrypt data in transit and at rest**

On a wider scale, although the number of data breaches of personal information has risen over the last few years across the globe, the volume of persons affected by said data breaches has begun to decrease[15]. Data protection regulation such as GDPR, alongside innovations in PETs and wider access controls within core systems, may be starting to have a positive effect on the overall risk associated with data breaches.

In parallel, innovations in digital identity, counter fraud technology and machine learning are continuing to reduce the threat of identity compromise, account takeover and other criminal acts enabled by the proliferation of breached personal data.

The DCMS Digital Identities and Attributes trust framework[16] is the clearest example as to how the combination of legislative and regulatory controls with modern technological and analytical techniques can revolutionise our collective approach to personal data, with the customer at the heart, and counter fraud as a core objective.

14 – https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/415469/Data_Sharing_for_the_Prevention_of_Fraud_-_Code_of_Practice__web_.pdf

15 – https://www.idtheftcenter.org/post/identity-theft-resource-center-reports-30-percent-decrease-in-data-breaches-so-far-in-2020/

16 – https://www.gov.uk/government/publications/uk-digital-identity-attributes-trust-framework-updated-version/uk-digital-identity-and-attributes-trust-framework-alpha-version-2

Conclusion – for now...

New and emerging PETs without doubt present tactical opportunities for fraud prevention and financial crime risk analysis. However, there is still work to be done to commoditise PETs, reduce costs, improve knowledge and best practice, and improve the quality of analytical outputs, before PETs become a credible alternative to current fraud/financial crime data intelligence platforms. This will be further supported as global compute power increases in line with Moore's Law[17].

Particular tactical use cases should focus on international data analytics, where data governance becomes obstructive to collaboration. As stated at the start of this paper, economic crime data analytics requires cross-border, cross-sector, and cross-organisation collaboration. Criminals, both opportunistic and organised, operate without borders and without restrictions.

PETs will help play a part in improving overall outcomes but in their current phase of development may not represent a silver bullet in 2022.

In terms of FATF's stated need for improved data sharing to disrupt economic crime, there are clearly lessons to be learned from well-established existing data sharing infrastructure/consortia that will provide significant operational benefits in economic crime disruption in the near term. PETs may be seen as an expensive distraction for the time being, but this could change as the market matures.

QUESTIONS?

Contact the Author:

Chris Lewis – Head of Solutions – Synectics Solutions

Email: **chris.lewis@synectics-solutions.com**

LinkedIn: **www.linkedin.com/in/chrisjlewis93/**

17 – https://www.umsl.edu/~siegelj/information_theory/projects/Bajramovic/www.umsl.edu/_abdcf/Cs4890/link1.html#:~:text=Moores%20law%20is%20said%20to,periodic%20increases%20in%20computing%20power