Enhancing Training and Technology Adoption in Terrorism Financing Investigations Through Gamification

Francesco Zola[1], Lander Segurola[1], Erin King[2], Martin Mullins[2] and Raul Orduna[1]
[1] Vicomtech Foundation, Basque Research and Technology Alliance (BRTA); Donostia; Spain
[2] University of Limerick; Castletroy; Ireland

fzola@vicomtech.org
lsegurola@vicomtech.org
erin.king@ul.ie
martin.mullins@ul.ie
rorduna@vicomtech.org

Abstract: The purpose of this publication is to present the methodology followed in the European project Anti-FinTer for training Law Enforcement Agencies (LEAs) and Financial Investigation Units (FIUs) in using emergent technologies to reveal financing activities of terrorism. The study presents, compares, and discusses the results gathered from three Capture-the-Flag events which involved LEAs and FIU officers. Designing curricula and training programs for improving terrorist financing investigations is challenging due to this domain's intricate and rapidly evolving nature and the multi-disciplinary knowledge needed. Furthermore, new tools based on novel paradigms, such as Artificial Intelligence and Big Data, are involved in terrorist financing investigations. However, they are too often unnecessarily complex and hard to use. These characteristics often limit law enforcement and end-users' engagement level and expertise in these technologies. For this reason, in this work, we describe an approach using gamification techniques to enhance technology and knowledge transfer for terrorist financing investigations. In fact, designing and implementing realistic and interactive challenges makes it possible to speed up the learning process, increase officers' expertise in using new technologies and improve their readiness. At the same time, this approach allows technical partners to gather end-user needs and facilitate development/validation cycles. This methodology has been validated in three pilots: one held in Madrid in 2022, a second in The Hague in 2023 and a final one in Vienna in 2023. In these pilots, law enforcement personnel were challenged in addressing tasks related to fighting financing terrorism activities through the dark Dark Web, crypto-assets or new payment systems. Results showed an increasing engagement, motivation, and knowledge in the participants.

Keywords: Gamifying, Hackathon, Counterterrorism, Law Enforcement Agencies, Cryptocurrencies, Training

1. Introduction

Tackling terrorist financing through investigation, prosecution, and prevention is a worldwide issue that extends beyond Europe. Every day, terrorists find new channels to communicate, campaign and finance their activities. For example, as reported by EUROPOL in the IOCTA report (Europol, 2021), the two main trends are crowdfunding campaigns and generating market revenue. In the first case, their *modus operandi* is straightforward: they raise a crowdfunding campaign to gather funds for their activities. In the second case, they try to generate revenue by selling extremist versions of common products or merchandise (such as Nazi-related items, ISIS promotional materials, etc.), as well as other legal and illegal goods (like counterfeit products, firearms, explosives, everyday items, etc.) to the public or other extremists/terrorists. In both cases, to maintain anonymity, they often employ a combination of cryptocurrencies and markets in darknet technologies (Europol, 2022).

To tackle these needs and combat cybercrime, new paradigms, such as Artificial Intelligence (AI) and Big Data, are being used alongside conventional software to create novel investigation tools (Maher, 2017). However, these tools typically include multiple steps for collecting, processing, analysing and visualising information related to financial data (e.g., transactions, electronic invoices, etc.) and correlating them with context data extracted from social media analysis, forums, phishing acts, etc. (Kilger & Choo, 2022). As a result, Law Enforcement Officers (LEOs) and other end-users may be deterred from using these technologies (Klingberg, 2022).

Designing curricula and training programs for improving terrorist financing investigations is challenging due to its intricate and rapidly evolving nature and also the different number of domains (and so tools) involved, from crypto finance to financial regulation, dark web structure, crypto ecosystem, etc. In that sense, a more practical approach can be a good solution to speed up the learning process and increase the readiness of the officers.

For this reason, in this paper, we present the deployment proposed in the Anti-FinTer (AFT) project (Anti-FinTer, 2022), in which traditional teaching techniques like lectures with moderated virtual learning environment (VLE),

workshops and exercises are combined with gamification techniques (hackathons) for facilitating interaction and engagement between participants. The aim of the AFT project is to train Law Enforcement Agencies (LEAs) and Financial Investigation Units (FIUs) to enhance their ability to use emergent technologies and complex data pipelines to reveal financing activities of terrorism. In this way, it will be possible to increase Europe's ability to use novel tools for investigating terrorist financing and promote EU technical and strategical sovereignty. The AFT project exploits four tools that have been developed in previous EU projects, such as Graphsense (Haslhofer, et al., 2021) for virtual assets analytics, the Visual Analytics tool for forensic image processing, Ordainsare as a transaction anomaly detector based on the model presented in (Zola, et al., 2019) and the Dark Web Monitor (CFLW, 2023) for analysing the darknet content.

Although the AFT project applies different teaching, learning, and training techniques, in this paper, our analysis is concentrated only on the latter. In particular, hackathon events are organized as training elements. These events are designed as Capture-the-Flag (CtF) exercises (Boopathi, 2015) that allow participants to learn effectively how to use new AI tools in their day-to-day work for revealing terrorist financing activities.

In this study, we present the general results obtained in three hackathon events, the first held in Madrid in Sept. 2022, the second held in The Hague in May 2023, and the third – and last - held in Vienna in Dec. 2023. The analysis performed after each event was used to identify limitations and organisational weaknesses that were addressed in preparation for the next hackathon event, allowing us to make improvements and provide a more professional service. The final results indicate satisfactory engagement among the participants; indeed, attendees achieved complete autonomy in using the AFT tools for realistic operations.

2. Preliminaries

2.1 End-user Profile

Participants are, in general, experienced practitioners in professions/industries that are exposed to terrorism financing, such as agents of LEAs or FIUs, and all arrive with a preconceived concept of the topic. In line with the expected learning outcomes of the project, we incorporate a central objective of understanding foundational and prevailing knowledge associated with terrorism financing and adjacent issues while providing participants with the autonomy to engage with learning materials they feel best suit their needs. The andragogical response to the training structure leverages the learning opportunities provided by these activities while also appreciating the diverse skill sets that the project participants possess (Harkin, 2022). In this sense, the main objectives of the gamification structure can be summarised as follows:

- Appreciate the requirements of the adult participants;
- Provide alternative routes of learning within the materials;
- Generate a learning environment rooted in experiential learning;
- Provide knowledge that can be perceived as immediately applicable or useful;
- Create a learning environment that can leverage the insights of training participants.

2.2 Gamification

Gamification techniques often make tasks or processes more engaging and enjoyable, encouraging participation, learning, or specific behaviours. In fact, leveraging the psychological and motivational aspects of games aims to increase user involvement, motivation, and achievement of goals. One of the strategies mainly used in information security contests is called Capture-the-Flag (CtF). CtFs are competitions where participants are called to address tasks and challenges to conquer flags. These flags can represent text, images, snippets of code, or a set of actions. Participants can work separately or in teams. In both cases, they are pitted against each other, testing their security skill.

At a high level, there are many types of cybersecurity competitions and platforms for managing them. In this sense, defining a finite set of CtF strategies is also difficult. Several studies (ENISA, 2021), (Švábenskỳ, et al., 2021) define two main CtF strategies: *Attack/Defend* and *Jeopardy-style*.

Attack/Defend is a type of format that involves an interactive competition game that requires at least two participants (Švábenskỳ, et al., 2021) divided into two teams: the attacker team (red) and the defender team (blue). The red task is to detect and exploit blues' vulnerabilities to conquer the flag. On the other hand, the blue team must resist and mitigate the reds' attacks by applying countermeasures. In this scenario, when the clock runs out, the end can result in getting the flag (red wins) or retaining the flag (blue wins). *Jeopardy-style* is a

format based on a challenge-based competition like the actual Jeopardy game with different categories and point values. The game consists of earning as many points as possible before the clock runs out. More than two users can be involved, and each starts by choosing a challenge from the board. When they find the solution to the chosen challenge (find the required flag), they submit it to the scoring system for evaluation. If the flag is correct, the team's score is updated, and the system allows the team to move on to the next challenge on the board.

The lack of attack/defend team specialization, the presence of multiple users, and especially the structured learning goals of the AFT project led us to choose the *Jeopardy-style* format to create a competitive environment and engage multiple users simultaneously.

2.3 Related Works

The CtF strategy has resulted in wide success in terms of introducing and learning cybersecurity-related concepts (Švábenskỳ, et al., 2021), but also motivating continued learning after the exercise (McDaniel, et al., 2016). For example, in (Huang, et al., 2011), CtF is used for solving as a differential game, whereas in (Eagle & Clark, 2004), this strategy is used to educate students to act as *crackers* and find new vulnerabilities in existing systems (data, files, devices, etc.). In (Werther et al., 2011), a CtF event based on a web server and application security is presented. In (Prinetto et al., 2020), authors propose a formal definition and a taxonomy for hardware-based CtF challenges. Although widely utilised in cybersecurity, there are only a limited number of scientific contributions available regarding the application of CtF to training LEAs. More specifically, the most used training strategies are based on teaching theoretical concepts and demonstrating the tools themselves without leaving the user to really use them in realistic operation. This happens in different EU projects such as (i-LEAD, 2017) project, where specific training sessions are organised to disseminate project technologies for tackling cybercrime and performing forensics investigations, or in (CYCLOPES, 2021), where events for training participants on specific Digital Forensics tools are organised. However, these events are planned like panels where LEA personnel and other stakeholders present their experiences and share their analysis and views on specific cybercrime topics without performing a real and practical task. Similarly, Joint Live Exercises are explored in (CTC, 2023) project to provide theoretical and practical knowledge about counter-terrorism financing and emerging terrorism financing risks. They are essentially showrooms for demonstrating the utility of the project platform. In the (DANTE, 2018) project, similarly structured training activities are organised to enhance knowledge on detecting and analysing terrorist-related content and financing activities. In this case, during the event, participants had the chance to use a demo version of the deployed tools, however, with limited data and without a defined scope. It is in the (ASGARD, 2016) project where training strategies based on CtF exercises are effectively exploited to speed up the LEAs training. However, the project aims to train the participants in using easily configurable and deployable tools, and for this reason, specific tasks are implemented, lacking their concrete application in real investigation.

Inspired by these previous works, we propose to use gamification strategies for training LEAs and FIU officers to use novel AI tools for terrorist financing investigations. More specifically, we propose a hands-on approach, which directly allows participants to use the AFT tools, leveraging the full spectrum of data they have available until the date of the events. Our methodology entails crafting challenges created from real-world use cases that help participants not only delve into tool functionalities but also uncover the tangible benefits they offer for their daily tasks. This immersive approach ensures that participants not only grasp the tools' capabilities but also develop the autonomy to utilise them effectively on their own.

3. Hackathon

3.1 Learning Requirements

Terrorism Financing and its associated issues are complex and ever-evolving topics that require intricate insight and knowledge across a broad range of subjects. These include political science, as it pertains to terrorist organisations; financial services and regulation around the use of cryptocurrency and fiat (real) money; technical knowledge on dark-web surveillance; policing; and financial/digital forensics. In particular, the focus of AFT is restricted to the dark web and crypto assets as facilitators of licit and illicit activities traceable to terrorist groups.

With these restrictions, it is important to leverage the prior knowledge and the subject area of interest expressed by the hackathon participants since each one has a diverse background and unique perspectives and needs. So, in that sense, the training must consider this fragmented scenario and design accessible and easy-to-solve tasks

that allow for homogenizing participants' theoretical and practical knowledge. Once this aim is achieved, introducing case study examples of terrorism financing issues can lead to practical, ready-to-apply knowledge. In fact, they are particularly effective in enhancing learning outcomes when real-world problems remain unresolved and ill-structured (Barrows, 2022). For this reason, the training approach introduced in the AFT project proposes using three strategies in the challenge definition: *tool-centric*, *category-oriented*, and *tool-free* (more information about them in Section 3.3).

3.2 Gamification Strategy

In the AFT project, CtF competitions based on a *Jeopardy-style* format serve as a dynamic and engaging gamification strategy. By adopting this approach, we aim to create an environment that fosters active participation and skill development among participants. All the challenges are aligned for discovering financing terrorism activities, which can include money laundering and fraud operations through the dark Dark Web, using crypto-assets or new payment systems and darknet marketplaces. More specifically, each challenge within the CtF competition is meticulously designed by the tool owners to assess and enhance participants' proficiency in utilizing their tools and, at the same time, to validate the functionality of these tools in real-world scenarios, demonstrating their efficacy in uncovering crucial hints and traces relevant to terrorism financing investigations. By highlighting the tangible benefits of these tools, they aim to inspire confidence and enthusiasm among participants, motivating them to further explore and utilize these innovative solutions in their professional endeavours.

On the other hand, a point-based system is employed to add an element of competition and motivation. In this sense, each challenge is assigned a score based on its complexity and difficulty level. Participants earn points for successfully completing tasks, with higher scores awarded for more challenging actions. Throughout the competition, participants accumulate points, and the individual with the highest score at the end is declared the winner. This point-based system not only incentivizes active participation but also fosters a sense of competitiveness among participants. Furthermore, to maintain competitiveness and better evaluate the engagement level of the participants, an individual-focused approach is used, with each team composed of just one participant.

A CtF managing platform called Facebook Capture-the-Flag (FBCtF) is employed to host the challenges. This platform allows users to define challenges on different levels, i.e., *Quiz, Flag* and *Basis*, according to the chosen strategy. In particular, for the Jeopardy-style used in AFT events, Flag level is used. This kind of challenge allows administrators to define a title, descriptions, category for the task, attachments (if applicable), the expected correct answer, the number of points obtained, a hint to avoid stuck users and a penalty for when the hint is used. Furthermore, the platform provides a management control where administrators can easily add new users to the platform, monitor their score and intent (game logs), and set other configurations.

3.3 Hackathon Format

As already introduced in Section 3.1, all along the AFT project, three different strategies are drawn and deployed to create hackathon challenges and to guide the users through the tools. More specifically, we started the project with a more structured and guided approach (Madrid), then proceeded to lessen constraints and limitations about the tools (The Hague) to finally enable participants to attain full independence in choosing the appropriate tools for the right tasks (Vienna). This dynamic methodology represents the primary innovation distinguishing AFT hackathons from other events. Indeed, this adaptive strategy enables alignment with the end user's evolving needs and progress, creating a continuously fresh and challenging environment. Furthermore, gradually releasing constraints helps the end user become proficient in using the tools for their day-to-day duties.

As shown in Figure 1, all three hackathon events shared the first *Introduction session* (30 minutes), in which organisers started with a welcome to the participants and explained to them the AFT goals and how the CtF works. However, to improve the learning process, each hackathon event follows a different strategy in the challenge definition, so they diverge from the rest of the agenda. Indeed, the Madrid event (2022) followed a *tool-centric* strategy, The Hague event (2023) was based on a *category-oriented* strategy, while the Vienna event (2023) used a *tool-free* approach. The change in the strategy influenced the number of planned sessions and the overall duration of the events.

The *tool-centric (first hackathon)* strategy implemented specific sessions (numbers 2, 3, and 4 in Figure 1) for testing each AFT tool separately. In these sessions, *easy* challenges related to the designated tools were

provided, aiming to assist users in becoming familiar with the tools, learning about basic functionalities, and evaluating tool usability. These tasks were self-explanatory to guide participants and prevent them from encountering obstacles. Thus, they needed to provide enough context for the investigation, be clear about the action to be performed and indicate the exact pattern to follow for writing the answer, avoiding grammatical errors. An example is shown as entry A in Table 1. In the fifth session of the first hackathon, participants were tested in more complex challenges that required deep domain knowledge as well as advanced expertise in utilising the AFT tools, as shown by entries B and C in Table 1. Each session lasted about 60 minutes, except for session number 3, where two tools were tested at the same time. A total number of 55 challenges were implemented in this hackathon.

The *category-oriented (second hackathon)* strategy planned an initial session (number 2 in Figure 1) to aid both new and returning participants to familiarise themselves with the tools and gain basic insights into their functionalities, as well as happened in sessions numbers 2, 3, and 4 of the Madrid events. Thereby, tasks were similar to entries A in Table 1. However, in this case, although the challenges contain the names of the tools to be used, they were not split within specific sessions but just put all together to speed up the learning process. Then, in session number 3 (Figure 1), the challenges were designed to tackle a real-world use case (UC) involving terrorism financing through a specific darknet market called *Luckp47*. This market is a niche market accessible using the Tor network, and it is claimed to belong to a paramilitary organisation (Jiang et al., 2021). In this darknet market, it is possible to buy different firearms, such as handguns, rifles, kalashnikov but also explosives, fake-ID and many other illicit goods using Bitcoin. Examples of tasks related to this UC is the entry D in Table 1. Again, the tool to be used was indicated in the title of each challenge. The tools were used separately, but each of them empowered users to accumulate evidence, facilitating their progress and comprehension through real investigation. The challenge-oriented ended with session number 4 (Figure 1), in which participants were asked to use the acquired knowledge to address more complex scenarios, which could also involve the usage of more than one tool at once (similar to entries B and C in Table 1). In this session, the idea was to show the benefits and limitations of the tools in scenarios not covered by previous sessions. On this occasion, each session required considerably more time to complete, and as a consequence of this approach, the second event extended beyond 5 hours (330 minutes). A total number of 78 challenges were available in this hackathon.

First Hackathon			Second Hackathon			Third Hackathon		
#	**Topic**	**Duration**	**#**	**Topic**	**Duration**	**#**	**Topic**	**Duration**
1	Introduction	30’	**1**	Introduction	30’	**1**	Introduction	30’
2	Dark Web Monitor	60’	**2**	Easy Challenges	90’	**2**	Easy Challenges	60’
3	Visual Analytics and Ordainsare	90’	**3**	Luckp47 UC	120’	**3**	Luckp47 UC	60’
4	GraphSense	60’	**4**	Hard Challenges	90’	**4**	Multiple UCs	60’
5	Medium/Hard Challenges	60’	**5**	Award Ceremony	-	**5**	Mobile Challenges	60’
6	Award Ceremony	-	**-**	-	-	**6**	Award Ceremony	-
Total		**300’**	**Total**		**330’**	**Total**		**270’**

Figure 1: Agenda first, second and third hackathon events.

Table 1: Examples of challenges for each category of Madrid, The Hague, and Vienna events.

Id	Tool	Category	Description/Context	Task
A	Visual Analytics	Easy	While browsing the existing dataset, you encounter an image depicting a gun over an orange background (see attachment).	**What is the prediction accuracy of the Top Concept (e.g. Glock)?**
B	GraphSense	Hard	Look at the US DoJ Statement of Facts in the Bitfinex arrest: *https://www.justice.gov/opa/pressrelease/file/1470186/download.*	**Based on the diagram on page 11, use GraphSense to determine the identity (the label/tag) of VCE4**.
C	Ordainsare	Hard	The attached file contains several addresses labelled as AlQueda activities.	**How many addresses are in the .csv? How many of them are correctly classified by Ordainsare? Among the classified, which is the most predominant LAST behaviour?**

Id	Tool	Category	Description/Context	Task
D	GraphSense	Luckp47 UC	Besides finding out who controls the funds, we may also be interested in learning more about the people who fund Luckp47.	**How many transactions were done directly from coinbase.com to the cluster containing the address "329NN882qvm69….LcXsh"?**
E	-	Multiple UCs	You have come across the Guns"R"Us webshop.	**Which exchange did the seller use to cash out?**
F	-	Multiple UCs	The BOAK has been known for attacking Russian military commissariats and telecommunication since 2022. Gathering information about this group, 2 Bitcoin addresses can be found.	**The first usage for one of the addresses (a1) was on 15/06/2020 at 01:52:02, and for the second one (a2) was on 29/09/2021 at 04:44:24. Which is the total amount that a1 has received? And a2?**
G	-	Mobile	Following some criminal investigations in the dark web, several suspicious addresses encoded in QR format are found.	**Which behaviour showed the address in its third appearance, and what classification score did it get?**

In the third hackathon, a *free-tool* strategy was implemented in order to remove constraints during the learning process. However, the first sessions, i.e., numbers 2 and 3 (Figure 1), were practically the same as sessions 2 and 3 of The Hague event, but with a reduced duration. In fact, these two sessions were used to assist participants in refreshing their tool knowledge and acquiring fundamental domain understanding. In session 4, new realistic challenges were implemented based on Guns'R'us and Arms Complex, two shops that allow the use of cryptos for buying weapons (similar to *Luckp47*). Users were asked to analyse the transactions of the addresses related to these two shops. At the same time, tasks related to "*Combat Organization of Anarcho-Communists*" (or *BOAK*) UC were implemented. *BOAK* is a militant anarcho-communist organization in Eastern Europe (Alexey Rozhkov, 2023), which often asks for donations for their purposes (CrimethInc., 2022). They accept Bitcoin donations, and several addresses are directly linked to known Exchanges. Nevertheless, following the free-tool strategy in the new task definitions, there was no indication of the tools to be used for completing the investigation, as shown by entries E and F in Table 1. The same approach was followed for creating the challenges of session 5, in which tasks were designed to show the tools' responsiveness and ability to adapt themselves to tablets and smartphones. These challenges required the usage of the device camera for taking pictures, converting images to text, scanning QR codes, etc, as shown by the entry G in Table 1. On this occasion, following user feedback gathered in the previous events, the overall duration of the hackathon was reduced to 270 minutes, and each session lasted 60 minutes. An overall number of 113 challenges were available in this hackathon.

3.4 Evaluation Metrics

Two evaluation frameworks are used for collecting feedback from the hackathon participants, one based on the common state-of-the-art *(a) learning schemes* and the other based on *(b) objective metrics*. In the *(a)* case, a survey with 35 questions is designed. This questionnaire is created following two distinct learning schemes: the System Usability Scale (SUS) (Peres, et al., 2013) and the Kirkpatric Model (Smidt, et al., 2009). The SUS is composed of 10 standardised questions used to assess the usability of a wide range of systems. These questions are rated on a five-point scale ranging from *Strongly Disagree* (1) to *Strongly Agree* (5). The questions cover a variety of factors that contribute to the overall usability, such as ease of use, efficiency, learnability, and satisfaction. On the other hand, the survey is improved with 25 (five-point Likert scale and open) questions based on the Kirkpatric model. This model comprises four criteria levels: Reaction, Learning, Behaviour, and Results (Smidt, et al., 2009). In the *(b)* case, regarding objective metrics, data and logs are directly extracted from the FBCtF platform and analysed in order to evaluate the trends and statistics of each participant as well as their engagement level. This analysis gives us an overview of the difficulties and problems encountered by each participant, and, at the same time, it helps us to determine the engagement level at each step. Yet, the results allow us to fill the gap between the perception of trainees about their learning process and their actual performance. For the sake of simplicity and also for the relevance of this work, just the results gathered from the objective metrics are reported and discussed in the next section. In fact, the survey results do not directly influence the presentation of the methodology's strengths and weaknesses.

4. Validation

4.1 Results

In the Madrid event, 16 participants attended the hackathon, 15 were in The Hague event, and 13 were in Vienna. Figure 2 reports the distribution of the participants' expertise in the three events. The figures show that, in all the events, the number of officers that work directly on real investigations (operational level) overwhelms the number of strategic ones. In fact, operational officers represent about 80% of the attendees. These outcomes align with expectations, as the AFT objective is to showcase the effectiveness of innovative tools for detecting terrorism financing primarily intended for operational officers' usage. Furthermore, we always had more LEAs rather than FIU officers.

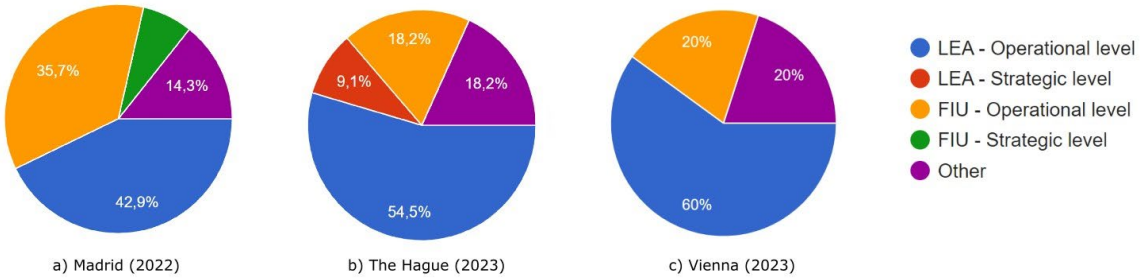

Figure 2: AFT users and external stakeholders' distribution in the first (a), second (b) and third (c) events.

The Madrid event showed a fairly consistent progress level among the participants, i.e., they were able to earn points without getting stuck for a long time. In fact, Figure 4 shows that four players (25%) achieved extraordinary scores of more than 2,000 points, and the other eight players (twelve in total, 75% of participants) achieved more than 1,425 points (black dotted line in Figure 4). Yet, two more players reached at least 1,000 points, whereas only two were below this last threshold. On average, the participants reached 1640 points out of the 2,850 available (red dotted line in Figure 4). On the other hand, Figure 3 shows that in The Hague event, the progress level was not as homogenous as in the first hackathon. In fact, despite the good feedback and comments gathered with the survey, only one player completed all the tasks and nearly attained the highest possible score, while six other additional players (less than 50% of the attendees) were able to reach at least half of the maximum score (black dotted line in Figure 3). In this event, the average score was 2,576 out of 4,350 available. Finally, as shown in Figure 5, more than 50% of the participants (7 out of 13) reached more than 3,000 points during the Vienna event. In particular, it is interesting that 11 out of 13 users, about 85%, reached scores higher than the average ones performed in The Hague and Madrid events and drawn as blue and black dotted lines in Figure 5, respectively. Therefore, the average score on this event was 3,496 points (green dotted line in Figure 5) out of 7,530 available (red dotted line). The participant's ability to achieve higher scores over successive events suggests a positive learning trajectory and potential knowledge retention. Furthermore, data indicate a progressive development of skills in using AI tools for terrorism financing investigations. Ultimately, the variation in performance levels shown in the different events suggests areas where the learning program could be refined for a more consistent and impactful training experience.

4.2 Discussion and Limitations

In general, the proposed methodology was shown to be suitable for training LEA and FIU officers and resulted in an improvement in the participant learning skills, as demonstrated by the increasing trend of average scores obtained in all the pilots. While one might question the significance of these improvements, considering that various challenges were previously addressed in past events, it's important to highlight that participants should recall how to use different tools even after extended periods (at least 6 months). At the same time, the improvements could also be related to a major number of challenges available in each newest hackathon. However, these concerns can be refuted considering that in the third hackathon, the overall event duration was reduced to 30 minutes and 60 minutes with respect to the first and second hackathons, respectively, and that information about the tool to be used in each task was removed. This confirms that, at this final stage, participants reached total autonomy and the highest level of tool and domain knowledge.

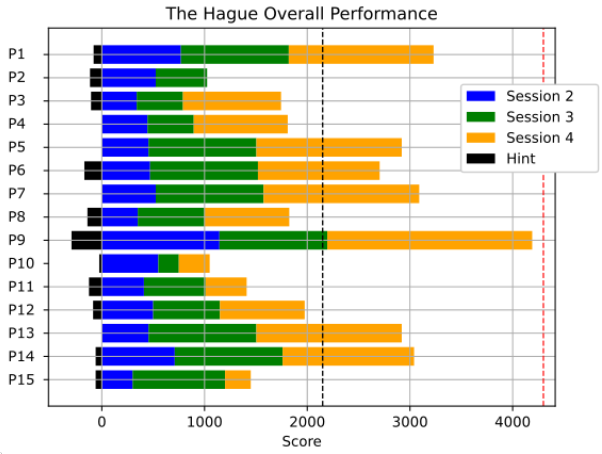

Figure 3: Participant score in the second hackathon

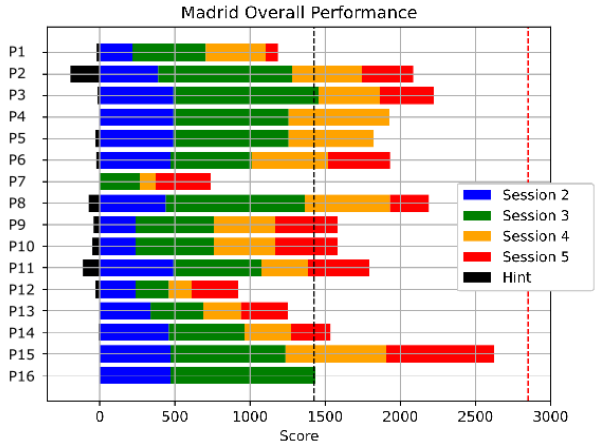

Figure 4: Participant score in the first hackathon

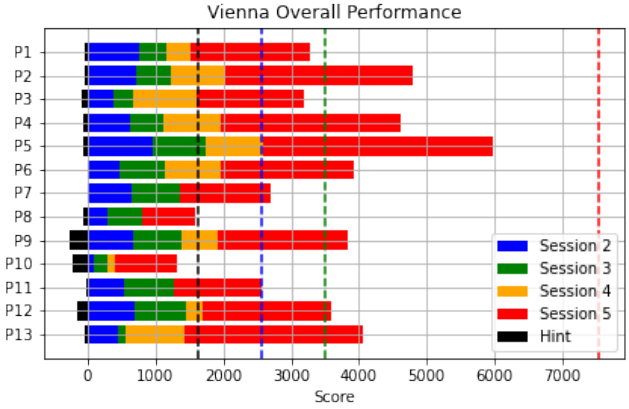

Figure 5: Participant score in the third hackathon

5. Conclusion

This paper describes a framework for training Law Enforcement Agencies and Financial Investigation Units to enhance their ability to use emergent technologies and complex pipelines to reveal financing activities of terrorism. The proposed approach used gamification techniques, such as Capture-the-Flag exercises, to engage the participants and to teach them how to use novel tools in their (realistic) investigations. The methodology was evaluated in three events held in Madrid (2022), The Hague (2023), and Vienna (2023) for training LEOs and FIUs in terrorism financing investigations. The overview of the results presented in Section 5 shown a satisfactory

level of engagement among the participants in all the hackathon events, as well as an incremental improvement in the acquired domain and technical knowledge.

Despite the positive results obtained, as a lesson learned, technical partners need to spend more time understanding the day-to-day needs of the agents to improve tool functionalities and adapt them to real investigations. At the same time, European Agencies and Commissions should keep working on trying to provide new ideas to foster cooperation and share knowledge and experiences on key challenges like new crypto-threats, financing trends and terrorism *modus operandi*. In this way, it will be possible to create a more homogeneous community, in which LEAs and FIUs can take inspiration for sharing experiences about new learning methodologies, realistic training/games and new tools/products useful for their investigations.

Acknowledgement

This work was partially funded by the European Union's Internal Security Fund — Police as a part of the Anti-FinTer project (grant agreement No. 101036262).

References

Alexey Rozhkov, C. A. D. P. F. C., 2023. *theanarchistlibrary.org.* [Online] Available at: https://en.wikipedia.org/wiki/Combat_Organization_of_Anarcho-Communists [Accessed 20 12 2023].

Anti-FinTer, 2022. *Versatile artificial intelligence investigative technologies for revealing online cross-border financing activities of terrorism.* [Online] Available at: https://anti-finter.eu/ [Accessed 20 12 2023].

ASGARD, 2016. *Analysis System for Gathered Raw Data.* [Online] Available at: https://www.asgard-project.eu/ [Accessed 20 12 2023].

Barrows, H., 2022. Is it truly possible to have such a thing as dPBL?. *Distance Education,* Volume 23, pp. 119-122.

Boopathi, K. S. S. a. B. A., 2015. Learning cyber security through gamification. *Indian Journal of Science and Technology,* Volume 8, pp. 642-649.

CFLW, 2023. *cflw.com.* [Online] Available at: https://cflw.com/dwm/ [Accessed 20 12 2023].

CrimethInc., 2022. *theanarchistlibrary.org.* [Online] Available at: https://theanarchistlibrary.org/library/crimethinc-russia-the-anarcho-communist-combat-organization [Accessed 20 12 2023].

CTC, 2023. *Cut The Cord.* [Online] Available at: https://ctc-project.eu/ [Accessed 20 12 2023].

CYCLOPES, 2021. *Fighting Cybercrime – Law Enforcement Practitioners' Network.* [Online] Available at: https://www.cyclopes-project.eu/ [Accessed 13 02 2024].

DANTE, 2018. *Detecting and Analysing Terrorist-Related Online Contents and Financing Activities.* [Online] Available at: https://www.h2020-dante.eu/ [Accessed 13 13 2024].

Eagle, C. & Clark, J. L., 2004. *Capture-the-flag: Learning computer security under fire,* s.l.: s.n.

ENISA, 2021. *All you need to know about Capture the Flag competitions.* s.l., s.n.

Europol, 2021. *Internet Organised Crime Threat Assessment (IOCTA).* s.l., s.n.

Europol, 2022. *European Union Terrorism Situation and Trend Report, Publications Office of the European Union.* s.l., s.n.

Harkin, D. a. W. C., 2022. Perceptions of police training needs in cyber-crime. *International Journal of Police Science & Management,* Volume 24, pp. 66-76.

Haslhofer, B., Stütz, R., Romiti, M. & King, R., 2021. GraphSense: A General-Purpose Cryptoasset Analytics Platform. *Arxiv pre-print.*

Huang, H., Ding, J., Zhang, W. & Tomlin, C. J., 2011. *A differential game approach to planning in adversarial scenarios: A case study on capture-the-flag.* s.l., s.n., p. 1451–1456.

i-LEAD, 2017. *innovation - Law Enforcement Agency's Dialogue.* [Online] Available at: https://cordis.europa.eu/project/id/740685 [Accessed 20 12 2023].

Jiang, C., Foye, J., Broadhurst, R. & Ball, M., 2021. Illicit firearms and other weapons on darknet markets. *Trends and Issues in Crime and Criminal Justice [electronic resource],* p. 1–20.

Kilger, M. & Choo, K.-K. R., 2022. *Do Dark Web and Cryptocurrencies Empower Cybercriminals?.* s.l., s.n., p. 277.

Klingberg, S., 2022. Countering Terrorism: Digital Policing of Open Source Intelligence and Social Media Using Artificial Intelligence. In: *Artificial Intelligence and National Security.* s.l.:Springer, p. 101–111.

Leune, K. & Petrilli Jr, S. J., 2017. *Using capture-the-flag to enhance the effectiveness of cybersecurity education.* s.l., s.n., p. 47–52.

Maher, D., 2017. Can artificial intelligence help in the war on cybercrime?. *Computer Fraud & Security,* Volume 2017, p. 7–9.

McDaniel, L., Talvi, E. & Hay, B., 2016. *Capture the flag as cyber security introduction.* s.l., s.n., p. 5479–5486.

Peres, S. C., Pham, T. & Phillips, R., 2013. *Validation of the system usability scale (SUS) SUS in the wild.* s.l., s.n., p. 192–196.

Prinetto, P., Roascio, G. & Varriale, A., 2020. *Hardware-based capture-the-flag challenges.* s.l., s.n., p. 1–8.

Smidt, A., Balandin, S., Sigafoos, J. & Reed, V. A., 2009. The Kirkpatrick model: A useful tool for evaluating training outcomes. *Journal of Intellectual and Developmental Disability,* Volume 34, p. 266–274.

Švábenskỳ, V., Čeleda, P., Vykopal, J. & Brišáková, S., 2021. Cybersecurity knowledge and skills taught in capture the flag challenges. *Computers & Security,* Volume 102, p. 102154.

Werther, J., Zhivich, M., Leek, T. & Zeldovich, N., 2011. *Experiences in cyber security education: The MIT Lincoln laboratory capture-the-flag exercise..* s.l., s.n.

Zola, F., Eguimendia, M., Bruse, J. L. & Urrutia, R. O., 2019. *Cascading machine learning to attack bitcoin anonymity.* s.l., s.n., p. 10–17.