UNCOVER: Development of an efficient steganalysis framework for uncovering hidden data in digital media

Vaila Leask, Rémi Cogranne, Dirk Borghys, Helena Bruyninckx

UNCOVER: Development of an efficient steganalysis framework for uncovering hidden data in digital media

Vaila Leask*
vaila.leask@mil.be
Royal Military Academy
Brussels, Belgium

Rémi Cogranne*
remi.cogranne@utt.fr
Université de Technologie Troyes
Troyes, France

Dirk Borghys
Royal Military Academy
Brussels, Belgium

Helena Bruyninckx
Royal Military Academy
Brussels, Belgium

ABSTRACT

This paper presents the general goals of Horizon 2020 project UNCOVER, whose overall purpose is to close the gap between academic work and operational needs in the fields of data-hiding. While digital data-hiding is a relatively new area of research, our motivation in this project has been rooted in the growing gap between the academic community and the operational needs of a "real-life" scenario of object inspection in order to *UNCOVER* the presence of data secretly hidden.

As well as an oversight into the structure of UNCOVER, our paper presents an empirical study on the impact of specifically training a detection method for a given data-hiding scheme, the so-called *Stego-Source Mismatch*, as an example of unexplored issues that raises important and mostly ignored consequences within the operational context the UNCOVER project targets.

KEYWORDS

Steganography, Steganalysis, Forensics, Cover-Source Mismatch, Image processing, Security, Data-hiding, Contest, H2020

ACM Reference Format:
Vaila Leask, Rémi Cogranne, Dirk Borghys, and Helena Bruyninckx. 2022. UNCOVER: Development of an efficient steganalysis framework for uncovering hidden data in digital media. In *Proceedings of Criminal Use of Information Hiding, Workshop of the 17th International Conference on Availability, Reliability and Security (ARES CUING'22).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

*Both authors contributed equally to this paper.

1 INTRODUCTION

Over recent years, steganography (the practice of concealing secret information within nonsecret media) has seen a rise in criminal use. At present, steganographic methods and technologies present a major challenge to Law Enforcement Agencies (LEAs) due to a lack of resources and procedures for investigations or structured operations. In order to carry out a full investigation into criminal and terrorist activities, LEAs currently use available (commercial) tools to detect hidden information in collected digital media. However, these tools detect only a limited number of hiding methods and lag a decade behind the scientific state-of-the-art. UNCOVER - a joint international initiative funded by the European Commission under the Horizon 2020 Research & Innovation program - aims to address these issues and further develop steganographic tools in order to establish a tailored toolkit for LEAs, as discussed in Section 3.

In this paper, we shall first provide a general background on steganography. This will then lead into a discussion about recent advances with state-of-the-art steganalysis and the LEAs current status in relation to the fields (Section 2). Following this, we shall discuss in more detail the structure, objectives and impacts of UNCOVER (Section 3) and present some early results obtained through the UNCOVER framework (Sections 4 & 5).

2 STEGANOGRAPHY & STEGANALYSIS

The term "steganography" comes from the Greek words *stegos* and *graphia*, meaning *covered writing*. Closely linked to cryptography (*hidden writing*), the difference between the two can be defined:

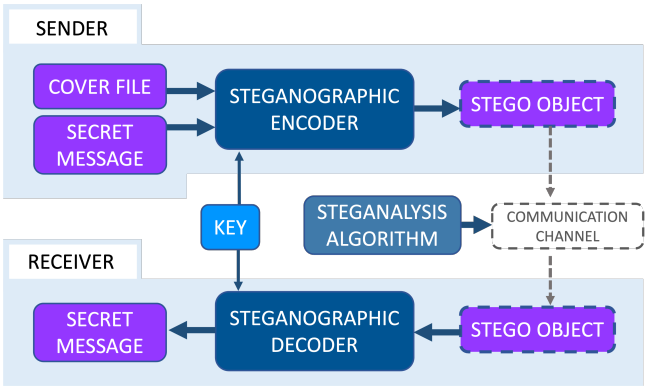

Figure 1: Basic steganographic model

Definition 2.1. In the context of retaining message confidentiality, **Cryptography** hides the *meaning* of a message, whereas **Steganography** hides the *presence* of a message.

Imagine a scenario where a sender has sent a secret message to a receiver encoded using a cryptographic scheme. Should a third party find the encrypted message, they would be able to deduce that the message was intentionally scrambled and thus this would raise suspicion. The secret message would then be open to being investigated and compromised. On the contrary, if the secret message is hidden within another, "innocent" file the intercepting party would not suspect the transfer of the message in the first place - therefore adding a considerable security feature whenever communication between parties could be considered compromising (note that the message can also be encrypted before hidden in the "innocent" media).

A basic steganographic model is depicted in Figure 1. A sender combines a cover file ("innocent" media) and a secret message with a secret key (used for embedding the message) in a steganographic encoder to create a stego-object. The cover file could be, for example, an image [1, 2], or a video [3], or nearly anything that can be digitally sent from one person to another [4]. The secret key is shared between sender and receiver by some external means and the stego-object is sent through the communication channel for the receiver to obtain. Once obtained, the receiver uses the key with a steganographic decoder to retrieve the secret message.

If an idealistic steganographic model were to be considered, the model would follow a cryptographic principle defined by 19th-century cryptographer, Kerckhoffs, stating: *the security of a cryptographic system should depend only on the secret key* [5]. Therefore meaning that a method of secretly encoding and transmitting information should remain secure even if everyone knows how it works. In fact, only the knowledge of the secret key will lead to a successful recovery of the secret message. In "real-life" applications steganographic models will unlikely follow Kerckhoffs' principle, which gives rise to the reverse process of steganography, *steganalysis*. Steganalysis aims at attacking the security of the steganographic scheme used to hide information by intercepting possible stego-objects through the communication channel between sender and receiver, as depicted in Figure 1. Generally speaking, the purpose of steganalysis is not to retrieve the message being sent, but rather simply confirm the existence of a secret message (this is due to the fact that the main goal of steganography lies in hiding the very existence of the secret message and its detection compromises the security of steganographic scheme). However, once detected, the stego-object can then be passed to the relevant party for a deeper analysis in order to attempt the retrieval of the embedding algorithm used, stego-secret key, message length or the message itself.

2.1 State-of-the-art steganalysis: data-hiding competitions

Recent advances in the field of steganalysis can be attributed to the launching of various data-hiding challenges. In this section, we provide insight into how the "real-life" test case scenario of steganalysis came to be explored.

2.1.1 ***BOWS: Break Our Watermarking System.*** Steganography and digital watermarking both aim to hide one piece of information (the *message*) inside another medium (the *cover*). Generally speaking, the difference between the two can be understood as [6]:

Definition 2.2. **Steganography** "undetectably" alters a cover: the message is an asset. The *cover* means to protect the *message*. **Watermarking** "imperceptibly" alters a work to embed a message about that work: the cover is an asset. In this context, the *message* means to protect the *cover*.

It should be noted, however, that there are other applications of watermarking, such as assessing the effectiveness of audio transmission [7]. One of the most recognizable forms of visible watermarking can be seen if one were, for example, to use a free-version of a commercially available editing software to alter the content (such as an image). In this situation, it is common practice that the software would allow one to edit the image with no issues, but then overlay their logo on the edited image once the user decides to save or download their final result (see Figure 2). The goal here is to protect the features available in the paid-version of the software and thus encourage the user to purchase their product. This process can also be applied invisibly to protect content: a good example of which is within the audio of cinema and blue-ray disks to protect the content from being pirated [8]. The decoder is embedded in the blue-ray player and if it detects the watermark while playing the disk it will conclude the disk is a pirated copy and thus stop the playback.

The European Network of Excellence in Cryptology (ECRYPT) supported the data-hiding community in launching two watermarking challenges, BOWS [9] and BOWS-2 [10] (*Break Our Watermarking System*), between 2005 and 2007. The purpose of the challenges was twofold: to assess the robustness and security of different watermarking systems, and to push research progress in the field overall. Both challenges were popular and saw the derivation of many novel approaches toward breaking watermarking systems. The success of BOWS and BOWS-2 henceforth inspired a drive towards assessing the robustness of steganographic systems, leading to a new challenge: BOSS (*Break Our Steganographic System*) [11].

2.1.2 ***BOSS: Break Our Steganographic System.*** One of the most successful approaches towards steganography in recent years is *content-adaptive steganography*, [12]:

(a) original image **(b) visibly watermarked image**

Figure 2: Example of visible watermarking on an image after using the free-version of a commercially available editing software

Definition 2.3. **Content-adaptive steganography** refers to a steganographic algorithm in which the payload is embedded while minimizing a relative distortion function. Therefore, this enables the steganographer to evaluate any distortion which may occur as a result of embedding changes.

An important motivation for the first steganalysis challenge, BOSS, was to evaluate the effectiveness of content-adaptive steganography for improving the empirical security of ensuing stego-media. To achieve this, in 2010 a new spatial-domain (meaning the pixel representation of an image) content-adaptive algorithm (HUGO - *Highly Undetectable steGO*) was designed for the creation of the competition's stego-objects.

BOSS advanced the field of steganalysis by forcing participants to deal with many new challenging problems [13–15]. While the competition was a success, it highlighted a significant problem for the practical applications of steganalyzers based on machine learning algorithms. The problem is known as *Cover-Source Mismatch* and will be discussed in section 2.2.

2.1.3 ***ALASKA.*** The BOSS challenge provided a large reference dataset for the steganalysis community and, while advances in the research community can be attributed to the use of this dataset, the specificity of content within the BOSS database was highlighted, [16]. BOSS bases were generated from RAW images captured with only 7 different cameras (in 2010, only high-end cameras allowed the exportation of RAW images) and those RAW files were developed into grayscale images, all using the same development pipeline - notably including a harsh resizing of the images to obtain a final image size of 512×512 pixels. This observation motivated the organisation of the ALASKA steganalysis challenges. Papers describing these challenges provide details regarding how academic research up until this point had been focused on image datasets with such specific features (grayscale, uncompressed, downscaled with a very high resizing factor, as with the BOSS database) [17, 18]. Additionally, steganalysis research works are often designed to benchmark steganography (as the BOSS challenge benchmarked the HUGO algorithm) which leads to the use of a worst-case scenario (following Kerckhoff's principle of cryptography, see Section 2.). This means that the steganalysist is provided with all information about the image generation process, the embedding rate, and the steganographic scheme - which is an unrealistic situation in "real-life" applications, such as operational forensic steganalysis. The goal of the ALASKA challenge was to move the application of steganalysis from a purely experimental environment to a more practical "real-life" environment. In particular, the image dataset was much larger (80, 000 images) and the images were developed in JPEG format using a combination of several different image processing algorithms to mimic what can be found in a more operational context. More details can be found in the aforementioned papers describing the challenges, [17, 18].

2.2 Cover-Source Mismatch

As mentioned in Section 2.1.2., the success of the BOSS challenge in 2011 highlighted a significant problem for practical applications of steganalyzers based on machine-learning: this problem is known as *Cover-Source Mismatch*. A thorough analysis of the Cover-Source Mismatch problem can be found in the papers[19, 20] which analyse the origin of image source heterogeneity and how this can affect the accuracy of steganalysis. However, for the readability of the present paper, we briefly recall the context of Cover-Source Mismatch (CSM) and some essential definitions below.

Definition 2.4. A **source** can be defined as an acquisition device (e.g. a camera), combined with a set of algorithms that generate cover contents such that for a given semantic content, the succession of acquisitions forms a stationary signal.

Definition 2.5. The term **Cover-Source Mismatch** (CSM) refers to the fact that when using two different sources for training a steganalysis method (usually based on a machine-learning algorithm), the learning outcome differs significantly while the set of embedding parameters (same algorithms, same embedding rate) and steganalysis method are the same.

One can note that CSM was already pointed out as one of the main barriers for the application of steganalysis under operational conditions in the review paper [21]. However, over almost one decade this problem has remained very seldom studied by the academic community.

2.3 Status in law enforcement

The modern world comes hand-in-hand with the rise in use of the internet. In parallel, a significant increase in the use of steganographic methods for criminal activities has been observed. This is attributed to the increased availability of steganographic tools, which have been made available as source code packages. Consequently, perpetrators can easily and selectively pick, adapt and combine information hiding tools for their criminal activities. An initial survey of the Criminal Use of Information Hiding (CUIng) initiative on the Europol Platform for Experts (EPE) revealed that evidence of steganography has been found in a wide variety of types of crime including child pornography [22], industrial espionage, criminal attacks on enterprises, credit card fraud & skimming, system intrusion, and backdoor injection & delivery methods [23].

Due to a lack of resources and procedures for structured operations, tackling steganographic technologies is a particularly challenging problem for LEAs - a problem which is heightened by the

increasing amount of digital evidence that LEAs and judicial partners have to handle. At present, LEAs use commercially-available tools to detect hidden information in digital media. These tools detect only a limited number of hiding methods, are slow, and offer no indication of confidence. Moreover, many commercial tools lag a decade behind the scientific state-of-the-art. The members of UNCOVER are committed to bridging these gaps and thus substantially increasing the technological autonomy of LEAs in the field of digital media steganalysis.

3 UNCOVER

The UNCOVER consortium consists of 22 multidisciplinary partners from 9 different European countries and is coordinated by the Royal Military Academy of Brussels, Belgium. The well-balanced consortium comprises of:

- LEAs
- Leading researchers from universities and other research institutions
- Partners in private and industrial sectors

3.1 Objectives & Impacts

With the goal of outperforming available steganalysis solutions in terms of performance, usability, operational needs, privacy protection, and chain-of-custody considerations, UNCOVER partners have joined forces to achieve the following eight objectives:

(1) Conduct a detailed analysis of the needs and requirements of LEAs for detecting and investigating steganography.
(2) Consolidate relevant information about existing steganographic tools and centralise this information in an intuitive database for LEAs.
(3) Improve existing methods for operational steganalysis in digital media workflows.
(4) Implement a flexible investigation platform.
(5) Demonstrate the steganographic detection capabilities with realistic test cases and scenarios delivered by the LEAs.
(6) Analyse the requirements in order to make the obtained results admissible in European court rules.
(7) Provide a comprehensive training program for LEAs and forensic institutes by providing in-house training.
(8) Validate the project results with practitioners, disseminate the outcomes, and prepare an exploitation plan.

A schematic overview of how UNCOVER will achieve these objectives is shown in Figure 3. By taking into account the requirements of LEAs at every step of this methodology, foreseen impacts of the previously defined objectives can be summarised:

- The LEAs and forensics institutes have the ability to detect and extract information hidden in different types of digital media.
- The UNCOVER tools will make the work of the LEAs and forensics institutes more efficient by speeding-up the processing time and reliability.
- UNCOVER will establish a network for cooperation, raising awareness, tracking progress, sharing information, working jointly, and training the staff.
- UNCOVER will contribute to the reduction or prevention of threats emanating from criminals and terrorists using steganography.
- UNCOVER will work towards a harmonisation of information formats at the international level, the improved cross-border acceptance and an exchange of court-proof evidence.

3.2 A General Issue: Fighting the Cover-Source Mismatch (CSM)

One of the main goals of the ALASKA challenge was to draw attention from the academic community to practical problems and scientific barriers that make the application of research works hardly usable in "real-life" scenarios [17, 18]. In this context, several facts were presented to show the difference between academic uses and practical needs, the most striking of which are the use of uncompressed and grayscale images, and the use of a reference dataset (namely BOSS [11]) in which all images are processed in the very same fashion (thus increasing so the CSM as reported in [16]).

The UNCOVER project aims at continuing this effort towards reducing the gap between academic works and practical applications, and the ALASKA challenge mostly focused on problems related with the CSM using content-adaptive state-of-the-art algorithms from the academic community. On the opposite, we have chosen, in this paper, to detail the almost unexplored problem of "*Stego-Source Mismatch*". For the sake of clarity, let us state that in the present study this term does not include the problem of recognizing the exact embedding which has been studied, see for instances [24, 25] and the references therein. In this paper the "*Stego-Source Mismatch*" is defined:

Definition 3.1. The **Stego-Source Mismatch** is the sensitivity of a steganalysis method with respect to the steganographic algorithm. In practice, this "*sensitivity*" is measured by how effectively a steganalysis method specifically designed for the detection of one specific steganography software is able to detect traces left by another, different tool.

Note that this problem can be also closely related to the so-called "universal steganalysis" which aims at detecting any embedding scheme and not targeting only one. Additionally, it should be noted that our study also aims at focusing on practical embedding software: meaning that (as all current state-of-the-art steganalysis methods are assessed against state-of-the-art embedding schemes) there is a considerable amount of software already in existence and readily available on the internet which are yet to be investigated.

This specific topic exemplifies the discrepancy between the research works and the needs of LEAs. On the one hand, LEAs will likely face stego-objects generated with steganography software downloaded from the internet however on the other hand, research in steganalysis focuses on the most advanced stego-tools. One could naively assume that when aiming to detect a state-of-the-art stego-tool such research should also detect stego-objects generated with less secure methods (such as those downloaded from the internet). However, it has been shown in the context of Cover-Source Mismatch that the transferability of steganlaysis remains a challenge and that, in fact, this approach of using state-of-the-art steganalysis on less-secure steganography methods is sub-optimal. The vast

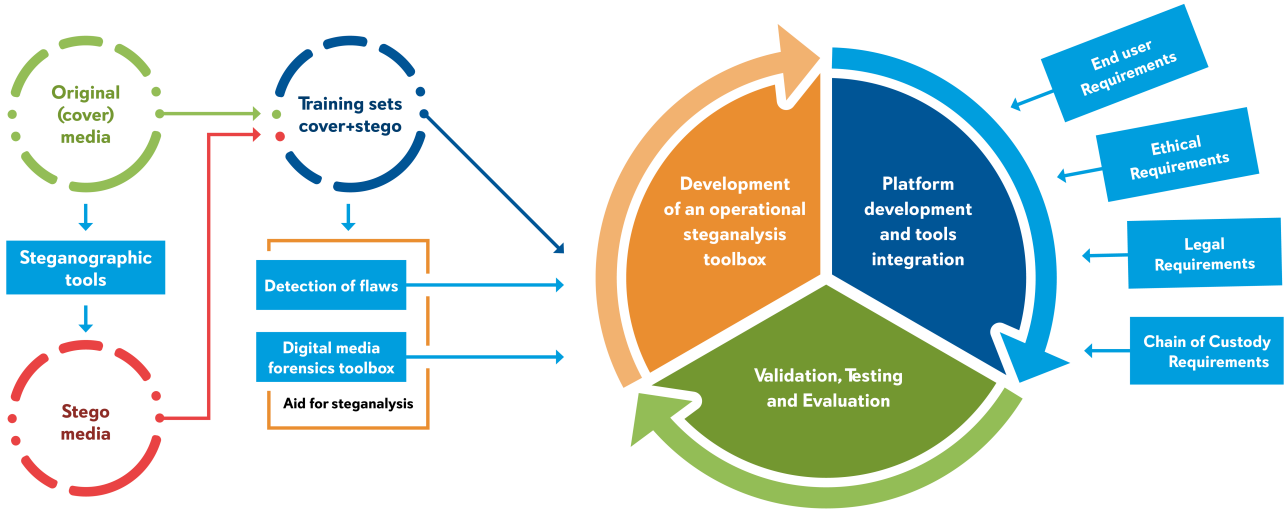

Figure 3: Schematic overview of UNCOVER

The methodological framework encompasses the following main steps: analysis of existing steganographic tools; development, training and theoretical validation of state-of-the-art detectors and tools; integration of tools into a user-friendly platform; field validation of UNCOVER solutions; and continuous feedback cycle.

majority of a stego-software available on the internet has been designed below the state-of-the-art benchmark and hence generates stego-objects leaving specific traces which can be easily detected; in this context, focusing on an extremely secure embedding scheme from the academic community would lead not to consider such traces despite their high practical interest.

4 EXPERIMENTAL SETUP: STUDY OF THE STEGO-SOURCE MISMATCH

As this is an empirical study, we shall start by presenting the experimental setup. The image dataset chosen was the ALASKA colour image dataset [17, 18] (which can be downloaded from kaggle or from the website https://alaska.utt.fr) and, for reproducibility, the embedding algorithms used for the ALASKA challenge were also chosen for this experiment (namely J-UNIWARD [26], UERD [27] and J-MiPOD [28] - note that the latter was improved in [29]). Due to the fact that the steganography algorithms / software selected for this experiment operate directly on the JPEG compressed images, we opted to use the version of the dataset compressed with different JPEG quality factors (using `libjpeg` version 8 used on our server within `pillow` package `python3`). Thus, a total of 120,000 images were used for training (60,000 cover-images and 60,000 stego-images) with another 20,000 for validation and 20,000 for testing.

Many different stego-software available on the internet were explored but it was concluded that only three would be used. For the purpose of studying the "Stego-Source" mismatch we needed to select embedding software that does not use any pre-processing, such as recompression or resizing, as these would create a strong "Cover-Source" Mismatch. To this end we focused on JPEG compressed images as those are, by far, the most widely used and hence would appear as the least suspicious cover. Furthermore, many software also re-compress the image during the embedding process, leading to two problems: first, when using a JPEG image as a cover, the recompression gives birth to a double compression which can be easily detected (raising suspicion about this image); second, when starting from an uncompressed image (note: the use of an uncompressed image as an original cover is not common in real life due to, for example, difficulty transferring the larger size of cover) the software may rely on a specific JPEG implementation and we do not want to detect this peculiar side-effect. As a final requirement, it was important to use software that can be used in command-line mode (and not only throughout the use of a graphical interface) so that we can generate large stego-image datasets.

Combining the aforementioned requirements and focusing on software both easily usable and available, three software were selected: *outguess*, *steghide* and *JPhide* (some of them are directly available from the main Linux repositories); note that *F5* also fit the requirements, but was unable to be used with our server because (even operating in the command-line mode) a graphical interface is required and thus prevents the generation of a large dataset on our server.

In order to comply with the usual practices in the academic community, the message length (or *payload*) depends on the number of non-zero AC DCT coefficients of each and every JPEG cover image. Note that for the dataset from the ALASKA2 steganalysis challenge, the payload was not constant (average of 0.4 bpnzAC, or 0.4 bits of secret message per non-zero AC coefficient). For the stego-software, a considerably smaller payload of 0.001 bpnzAC was used,

as this prevents us from falling into a trivial detection problem from the point-of-view of operational steganalysis. Additionally, three JPEG compression rates were used - defined by the standardized JPEG quantization matrix corresponding to quality factor (QF) 100, 95 and 75 respectively.

For the steganalysis itself, the Deep Learning model *Efficient-B3* [30] was shown to be particularly effective during the ALASKA2 steganalysis challenge and so was also selected for this experiment. We have used a curriculum learning technique which consists, for the application in steganalysis, iteratively training the network starting with a higher payload in order to ease the convergence of the training process. In all our experimentation, we have used the `pytorch` implementation of EfficientNet with `adamW` optimizer over NVIDIA RTX 3090 GPU, allowing us to use a batch size of 24 images; we started with a learning rate of 0.001 and a scheduler "`Reduce on plateau`" with reduction factor 0.5 and patience parameter 1 while the number of epochs is set to 15 (after curriculum during which we used only one single epoch).

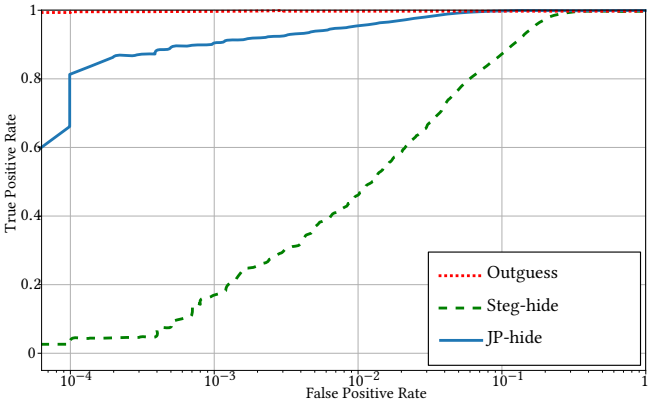

Figure 4: ROC Curve obtained with EfficientNet-B3 over image compressed at QF=75

5 RESULTS

As the main goal of the present paper is to empirically study the *Stego-Source Mismatch* with a special attention devoted to real embedding software, we go straight to the point by looking at Tables 1-2. The two tables provide the same results over images compressed with JPEG standard at $QF = 75$, table 1, and at $QF = 95$, table 2.

The tables present the total probability of error, under equal prior, usually denoted P_E ; the rows correspond to the embedding software used in the testing set and, on the opposite, the columns represent the embedding method used for the training set (including validation). To be more specific, we would like to emphasize that the code used a seeded-pseudo random number generator such that the training, validation and testing sets are always the same regardless of the steganographic embedding.

There are two obvious results one can conclude from Tables 1-2. First, looking at the "diagonal" leftmost part of those tables (with a light-grey background), when the embedding schemes match during the training and testing sets, the embedding software we have chosen are rather simple to detect, while those are amongst the most advanced readily available from the internet. More specifically, it seems that *Outguess* leaves some specific traces that a well-trained deep learning model can efficiently detect. Similarly, *JP-hide* also seems very easily detectable for low QF but surprisingly more secure for $QF = 95$. The third software, *Steghide*, is the most secure among the three embedding software we used; we would like to recall that for an easier comparison we used the same and very low embedding rate for all those embedding software of 0.001 bpnzAC (bits per non-zero AC coefficients) which actually correspond to a few dozens of bytes embedded in most of the case for $QF = 75$.

The second striking observation from Tables 1-2 is that the detection accuracy depends very much on the specific stego-source used during the training phase. Looking at the rows of those tables one can see that, when the training is carried out using a different embedding technique, the detection accuracy is very significantly reduced; in fact, in the vast majority of cases the Stego-Source Mismatch leads to a detection comparable to a random guess. Interestingly, one can note that this observation always holds true regardless of the testing algorithms. Using either the most-advanced embedding algorithms from the academic community (namely, J-MiPOD, J-UNIWARD or UERD) or using similar software for training does not seem to make the transferability of steganalysis straightforward detection or even doubtful.

To end with a more positive result, it is of interest to the authors to further investigate the detection results of real-life steganographic algorithms using EfficientNet-B3, which is among the state-of-the-art for steganalysis. The UNCOVER project focuses on practical applications of steganalysis for Law Enforcement Agencies and, while the Stego-Source Mismatch is an important aspect, the possibility to achieve a very reliable detection (with a very low false-positive rate) also constitutes another major barrier in this direction.

To this end, Figures 4-5 present the so-called ROC[1] curves, plotting the True-positive detection rate as a function of the False-positive rate. Note that for readability those figures are drawn using a logarithmic scale on the x-axis in order to feature low-positive rates. Clearly Figures 4-5 show that *Outguess* embedding software can be detected very reliably. This may be due to a specific signature, and more investigation to understand this phenomenon is needed. Similarly, *JPHide* can also be detected with high reliability with both $QF = 75$ and $QF = 95$ since the True-Positive rate remains as high 50% for a very low False-Positive rate of 10^{-4}. In practice, this seems very much acceptable as this means detecting "only" half of the stego-objects but almost never erroneously raising an alarm for a cover falsely classified as a stego. Note that in this specific case we used a total of $20,000$ images (cover and stego) which can explain large variation for False-positive rates as low as 10^{-4} as those actually correspond to a few samples. However, one can note that, on the opposite, the detection of *Steghide* with a deep learning method "as it" seems out of reach and that a specific training method focusing on low false-positive rates is badly needed in that case.

[1]ROC stands for Receiver Operational Characteristics ; this not so explicit name is due to the fact that such a plot was originally developed for operators of radar receivers.

		Embedding algorithm used for: Training phase					
		JPhide	Outguess	Steghide	J-MiPOD	UERD	J-UNIWARD
Testing phase	JPHide	2.37%	49.87%	48.24%	46.37%	48.26%	27.01%
	Outguess	49.13%	0.16%	42.31%	49.04%	29.73%	37.45%
	Steghide	49.69%	49.89%	10.55%	44.71%	49.72%	49.04%

Table 1: Empirical results on steganalysis efficiency (in P_E) depending on the training embedding algorithm for $QF = 75$.

		Embedding algorithm used for: Training phase					
		JPhide	Outguess	Steghide	J-MiPOD	UERD	J-UNIWARD
Testing phase	JPHide	15.56%	49.80%	49.84%	48.69%	49.47%	44.58%
	Outguess	48.57%	0.14%	41.08%	47.82%	23.46%	38.24%
	Steghide	49.80%	49.66%	22.76%	46.04%	49.83%	49.10%

Table 2: Empirical results on steganalysis efficiency (in P_E) depending on the training embedding algorithm for $QF = 95$.

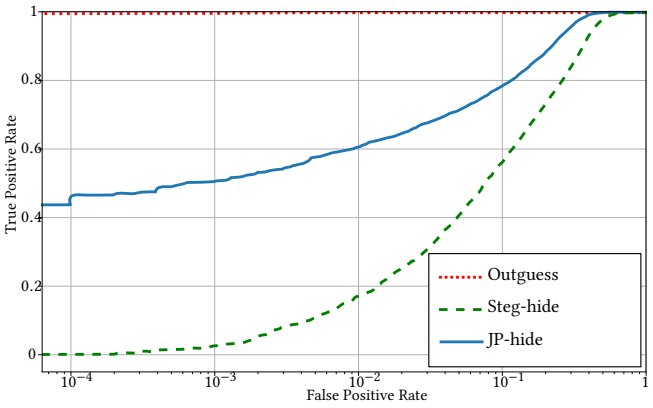

Figure 5: ROC Curve obtained with EfficientNet-B3 over image compressed at $QF = 95$

6 FUTURE WORK

The main goal of the experimental results provided in this paper is to show the reader that in the fields of steganography and steganalysis, many practical issues of major importance are seldom studied by the academic community. We have also explained that some of those major barriers can only be lifted with scientific advances and not only engineering work. This has been exemplified in the present paper by the problem of the extremely large heterogeneity of cover- and stego-objects one has to deal with in a practical situation while detection can heavily depend on each and every factor that gives birth to this massive diversity, some of those factors being not even clearly identified.

In addition, we have shown that the current steganalysis tools are extremely efficient and can be used directly for the detection of steganographic software one can find on the internet. However, it was also demonstrated that learning methods focusing on the reliability of the detection, in the sense of controllable and very-low false-positive rate, constitutes a major challenge that can only be addressed by novel scientific methods.

The results presented in the present paper are based on the ALASKA dataset [17, 18]. Additional works on extremely diverse dataset such as those that one can find on the internet is also required to confirm that an extremely reliable detector can be achieved in this operational context.

Within the UNCOVER project, we aim to focus on those often unexplored aspects of steganalysis from different points of view including understanding the source of media heterogeneity better, improving forensics analysis to classify media origin, investigating signatures left by specific embedding methods, and how to perform reliable detection in this complex environment.

7 CONCLUSIONS

This paper provided a general presentation of Horizon 2020 project UNCOVER, whose main goal is to move steganalysis research works closer to the needs of the operational context. We have sketched how the data-hiding competitions helped the community while also pushing to focus on specific cases which are not always very realistic.

In order to exemplify how much moving steganalysis into an operation context raises unanswered questions, this paper presents a short empirical study on Stego-Source Mismatch between embedding software available on the Internet. Our findings confirm that, on the one hand, such software is much easier to detect with high accuracy, *i.e.* with a very low false-positive rate. On the other hand, the so-called *universal steganalysis*, that is the detection of a wide range of data-hiding schemes, remains a challenging problem.

8 ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101021687.

REFERENCES

[1] Abbas Cheddad, Joan Condell, Kevin Curran, and Paul Mc Kevitt. Digital image steganography: Survey and analysis of current methods. *Journal signal processing, Volume 90, Issue*, 2010.
[2] H.G. Schaathun. *Machine Learning in Image Steganalysis.* IEEE Press. Wiley, 2012.
[3] Yunxia Liu, Shuyang Liu, Yonghao Wang, Hongguo Zhao, and Si Liu. Video steganography: A review. *Neurocomputing*, 335:238–250, 2019.
[4] Souvik Bhattacharyya, Dr. Indradip Banerjee, and Gautam Sanyal. A survey of steganography and steganalysis technique in image, text, audio and video as cover carrier. *Journal of Global Research in Computer Sciences*, 2:1–16, 2011.
[5] Auguste Kerckhoffs. La cryptographie militaire. *Journal des sciences militaires*, 4:5–38, 1883.
[6] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital Watermarking and Steganography.* Morgan Kaufmann, 2nd. edition, 2007.
[7] Gustavo M. Calixto, Alan C. B. Angeluci, Celso S. Kurashima, Roseli de Deus Lopes, and Marcelo K. Zuffo. Effectiveness analysis of audio watermark tags for iptv second screen applications and synchronization. In *2014*

International Telecommunications Symposium (ITS), pages 1–5, 2014.
[8] Ron van Schyndel. Watermark-based media annotation for blu-ray disks. In *OzeWAI (Australian Web Accessibility Initiative)*, 12 2009.
[9] A. Piva and M. Barni. The first BOWS contest (Break Our Watermarking System). In Edward J. Delp III and Ping Wah Wong, editors, *Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, pages 425 – 434. International Society for Optics and Photonics, SPIE, 2007.
[10] P. Bas and T. Furon. Bows-2, 2007.
[11] Patrick Bas, Tomáš Filler, and Tomáš Pevný. "break our steganographic system": The ins and outs of organizing boss. In Tomáš Filler, Tomáš Pevný, Scott Craver, and Andrew Ker, editors, *Information Hiding*, pages 59–70, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
[12] Vojtěch Holub et al. *Content Adaptive Steganography: Design and Detection*. Citeseer, 2014.
[13] Jessica Fridrich, Jan Kodovský, Vojtěch Holub, and Miroslav Goljan. Steganalysis of content-adaptive steganography in spatial domain. In Tomáš Filler, Tomáš Pevný, Scott Craver, and Andrew Ker, editors, *Information Hiding*, pages 102–117, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
[14] Jessica Fridrich, Jan Kodovský, Vojtěch Holub, and Miroslav Goljan. Breaking hugo – the process discovery. In Tomáš Filler, Tomáš Pevný, Scott Craver, and Andrew Ker, editors, *Information Hiding*, pages 85–101, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
[15] Gokhan Gul and Fatih Kurugollu. A new methodology in steganalysis: Breaking highly undetectable steganograpy (hugo). In Tomáš Filler, Tomáš Pevný, Scott Craver, and Andrew Ker, editors, *Information Hiding*, pages 71–84, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
[16] Vahid Sedighi, Jessica J. Fridrich, and Rémi Cogranne. Toss that bossbase, alice! In *Media Watermarking, Security, and Forensics*, Proc. IS&T, pages pp. 1–9, Feb 2016.
[17] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. The ALASKA Steganalysis Challenge: A First Step Towards Steganalysis "Into The Wild". In *ACM IH&MMSec (Information Hiding & Multimedia Security)*, ACM IH&MMSec (Information Hiding & Multimedia Security), Paris, France, July 2019.
[18] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. Alaskav2: Challenging academic research on steganalysis with realistic images. In *Information Forensics and Security (WIFS), IEEE 12th International Workshop on*, page 4, December 2020.
[19] Quentin Giboulot, Rémi Cogranne, and Patrick Bas. Steganalysis into the wild: How to define a source? In *Media Watermarking, Security, and Forensics*, Proc. IS&T, pages 318–1 – 318–12, Jan 2018.
[20] Quentin Giboulot, Rémi Cogranne, Dirk Borghys, and Patrick Bas. Effects and solutions of cover-source mismatch in image steganalysis. *Signal Processing: Image Communication*, 86:115888, 2020.
[21] Andrew D. Ker, Patrick Bas, Rainer Böhme, Rémi Cogranne, Scott Craver, Tomáš Filler, Jessica Fridrich, and Tomáš Pevný. Moving steganography and steganalysis from the laboratory into the real world. In *Proceedings of the first ACM workshop on Information hiding and multimedia security*, IH&MMSec '13, pages 45–58, New York, NY, USA, 2013. ACM.
[22] K. Cabaj, L. Caviglione, W. Mazurczyk, S. Wendzel, A. Woodward, and S. Zander. The new threats of information hiding: the road ahead, 2018.
[23] European Commission under the Horizon 2020 Research & Innovation program. Uncover, 2021.
[24] Tomas Pevny and Jessica Fridrich. Multiclass detector of current steganographic methods for jpeg format. *IEEE Transactions on Information Forensics and Security*, 3(4):635–650, 2008.
[25] Rémi Cogranne and Jessica Fridrich. Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory. *IEEE Transactions on Information Forensics and Security*, 10(12):2627–2642, 2015.
[26] Vojtěch Holub, Jessica Fridrich, and Tomàš Denemark. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security*, 2014(1):1–13, 2014.
[27] Linjie Guo, Jiangqun Ni, Wenkang Su, Chengpei Tang, and Yun-Qing Shi. Using statistical image model for jpeg steganography: uniform embedding revisited. *IEEE Transactions on Information Forensics and Security*, 10(12):2669–2680, 2015.
[28] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. Steganography by minimizing statistical detectability: The cases of jpeg and color images. In *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec'20, pages 161–167, New York, NY, USA, 2020. Association for Computing Machinery.
[29] Rémi Cogranne, Quentin Giboulot, and Patrick Bas. Efficient steganography in jpeg images by minimizing performance of optimal detector. *IEEE Transactions on Information Forensics and Security*, 17:1328–1343, 2022.
[30] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning (ICML)*, pages 6105–6114. PMLR, 2019.