CONSIDERATIONS ON THE BENCHMARKING OF MEDIA FORENSICS
Christian Kraetzer, Jana Dittmann
Otto-von-Guericke-University Magdeburg, Dept. of Computer Science
Magdeburg, Germany

ABSTRACT
In media forensics, most of the approaches are still far away from being mature enough for admission in court – the ultimate benchmark for any forensic method.

The intended contribution of this paper is to facilitate the understanding between the researchers active in this field and the society perspective on media forensics. The intended benefit of this is that it allows these two to better understand each other: On one hand, showing researchers working on the development of forensic methods a precise picture of the compliance requirements installed by the society (here, requirements for potential court admissibility); on the other hand, intending to help non-technicians to understand the challenges that researchers face.

To achieve this contribution, a selected prominent example (the photo response non-uniformity (PRNU) based digital camera forensics) is discussed here briefly in the context of the established Daubert criteria as reference imposed as admissibility threshold for legal proceedings.

Index Terms — Benchmarking, media forensics, court admissibility, Daubert criteria

1. INTRODUCTION
The motivation for performing media forensics (here, especially the verification of the authenticity and integrity of media objects such as digital images or audio data) is apparent in our modern society: an increasing number of data objects we encounter are already generated as digital objects. This fact is also true for data that is examined in forensic investigations. Here, traditional methods for forensic investigations on photos or audio tapes (see e.g. the summary presented in [1]) have to be replaced by adequate digital counterparts.

Regarding the issue of trust (i.e. value) in forensic procedures and techniques, questions arise in the society. This can be very well illustrated on the example of fingerprint analysis – one of the oldest and most respected forensic sciences: Since being established in the 19th century, it has been used worldwide to solve criminal cases by matching latent fingerprint traces found e.g. at crime scenes to samples acquired from suspects. But even an established practice like fingerprint analysis has to face new challenges in public debate. One extremely influential challenge to fingerprint analysis has been the amendment of the U.S. Federal Rules of Evidence (FRE; [9], [5]) in the year 2000 as a result of the two ground-breaking cases Dauber (Daubert v. Merrell Dow Pharmaceuticals, Inc. United States Court 509 U.S. 579, 1993) and Kumho Tire (Kumho Tire Co. v. Carmichael, 526 U.S. 137, 119 S. Ct. 1167, 1999). This amendment to the FRE led during a short time after its installment to several court cases, where the presiding judge refused to allow the admission of fingerprint analysis results in court on the basis of the fact that fingerprint experts, despite the fact that they are highly trained and certified specialists, are no scientists (see [2]). Up until today, the consideration of forensic compliance, e.g. to the Daubert standard or the so-called Daubert criteria (see section 2), has been grossly neglected in the field of media forensics. The main reason for this fact has to be sought in the current practice of implementing myriads of highly specialized approaches for very narrow application scenarios. This leads to a wide landscape of forensic tools and approaches, most of them promoted by only one person or one research group. Naturally, very few of these approaches will ever see the necessary field penetration and acceptance that would be for passing the hurdle set by the Daubert hearings.

Some specific application scenarios are beginning to see approaches that achieve some degree of generalizability, but approaches that span different application scenarios are still amiss in the field of media forensics.

We intend to facilitate the understanding between the research and the society perspectives on media forensics by summarizing some fundamental issues. On one hand, it shows IT-security researchers working on the development of forensic methods a precise picture of the compliance requirements installed by the society; on the other hand, it is intended to help non-technicians to understand the challenges that researchers in IT-security face.

The rest of this paper is structured as follows: Section 2 summarizes the Daubert standard (parts of the U.S. Federal Rules of Evidence (FRE) as well as the Daubert criteria) as performance indicators for forensic compliance. In section 3 a discussion on the maturity of one exemplary selected application scenario (the photo response non-uniformity (PRNU) based digital camera forensics) is presented. The section 4 summarizes the assessment made and presents a conclusion of the work presented.

2. IT-FORENSICS AND THE DAUBERT CRITERIA AS FORENSIC COMPLIANCE INDICATORS
According to a well-established definition (cf. [3]) given in [4], IT-forensics is: “The use of scientifically derived and proven methods [...] for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown
to be disruptive to planned operations.” For this paper we assume that requirements for “scientifically derived” and “proven” forensic conformity or compliance are considered to be imposed by the Daubert standard, a set of rules defined for U.S. federal level judicial matters in the 1990s by the United States of America Supreme Court based on the decisions in the Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S. Ct. 2786 (1993) or short ‘Daubert’ court cases. The Daubert standard or, more precisely, the requirements of the FRE rule 702 (7) and the Daubert criteria are verbally described requirements (rules) which leave room for the judicial interpretation required in any specific court case and it is widely regarded as a good (if not the best established) set of recommendations for judges on how to evaluate the usefulness of scientific (as well as non-scientific) expert testimony (see e.g. [6]).

2.1. Rules governing the interpretation of forensic results in (US federal level) courts

Without having sophisticated background in law to completely evaluate the whole set of legal challenges to the admissibility of traces or even evidence that is generated in this way, some basics on this matter have to be discussed here, because this admissibility would be the ultimate benchmark for every forensic method. All legal considerations presented here are based on freely available material concerning the U.S. legal system at the federal level. The Federal Rules of Evidence (FRE)1, define the framework within which evidence can be admitted into court. Even if these rules are in their original form only applicable on U.S. federal level, their concepts for handling forensic data have, to the best of the author’s knowledge2, influenced many other judicial systems worldwide.

In general, forensic results, like the ones considered within this paper, have to be interpreted by experts to the court. Using the terminology of U.S. jurisdiction, the trial judge acts as a form of ‘gatekeeper’, assuring that scientific expert testimony interpreting the findings of a forensic investigation truly proceeds from reliable (or scientific) knowledge. Considerations on relevance and reliability require the trial judge to ensure that the expert’s testimony is ‘relevant to the task at hand’ and that it rests ‘on a reliable foundation’. According to [2], the primary rules that are relevant for the presentation of forensic evidence in court (i.e. that apply to expert witnesses) in the FRE are FRE rule 702 ("Testimony by Experts") and FRE rule 703 ("Bases of Opinion Testimony by Experts"). Regarding the admissibility of an expert, the judge has to establish whether the following four points are met:

- **Qualification of a witness as expert**: First, a witness has to qualify as an expert. For the description of the involved process for U.S. legal system see e.g. [2] or [8]. The judge decides whether and to which extend the witness may offer opinion testimony as an expert.

- **Type of knowledge considered**: FRE rule 702 specifies different types of knowledge that an expert can offer. According to [2] the question is which kind of knowledge is considered is addressed by the U.S. supreme court in Kumho Tire. In this case the court states that the same evaluation criteria used in Daubert to determine whether testimony offered as scientific knowledge is reliable should also govern the admissibility of testimony for the “technical” and “other specialized knowledge” [7] types of knowledge. Therefore, distinguishing between science, applied science, technology, or experience-based expertise is not required.

- **Who is addressed by the expert**: Basically, two entities of expert have to be convinced. First, the judge, to get admitted in pre-trial hearings, and second the ‘fact finder’ (the "trier of fact" in FRE rule 702 [7], either a jury in normal cases or a judge in non-jury trials) at the trial itself.

- **Qualification**: Any expert has to testify upon the criteria "knowledge, skill, experience, training, or education" [7]. This information helps to decide whether an expert can be admitted and helps the ‘fact finder’ to assign corresponding weights to each expert’s testimony in the decision process.

The year 2011 version of FRE rule 702 ("Testimony by Experts") stated (see e.g. [7]; FRE as amended Apr. 26, 2011, (eff. Dec. 1, 2011) imposes the following four requirements as evaluation criteria for forensic investigations:

- FREC0: "the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue"  
- FREC1: The investigation (which leads to the corresponding expert testimony) is "based upon sufficient facts or data"  
- FREC2: The investigation is based upon "reliable principles and methods", preferably scientific methodology and knowledge  
- FREC3: The methods are applied "reliably to the facts of the case"

In the notes on FRE rule 702 published by the Legal Information Institute at Cornell Law School in December 2010 [10]) the interpretation of this rule is summarized as follows: “Rule 702 has been amended in response to Daubert [...], and to the many cases applying Daubert, including Kumho Tire [...]. In Daubert the Court charged trial judges with the responsibility of acting as gatekeepers to exclude unreliable expert testimony, [...].” The main result of this amendment, the so called Daubert hearings, is discussed in section 2.2 below.

The second important rule regarding expert testimony is FRE rule 703 ("Bases of an Expert’s Opinion Testimony"; see e.g. [9]), specifying two types of testimony experts are allowed to offer in court – first-hand knowledge, where an expert acts as a fact witness, testifying on

---

1 These rules are a ground-breaking law reform in U.S. federal law, specifying strict general rules, approved by the U.S. Supreme Court and the Congress passed the FRE in 1975. The FRE became effective for all U.S. federal courts on July 1st, 1975.

2 Important notice: The authors have absolutely no legal training. All legal considerations made within this paper are therefore layman's interpretation of freely available material, which are made to the best of our knowledge. Our intention is to motivate the research community to look further in this direction – the paper is not intended to be used in a legal proceeding. Irrespective of our intention, if the content of this paper is intended to be used in any legal proceedings, the reader must consult appropriate legal counsel for the corresponding jurisdiction.
observations made in examining evidence (as important component of this rule within this paper) and the testimony based on reports or examinations made by others.

The remaining FRE regarding opinions and expert testimony (rule 701 “Opinion Testimony by Lay Witnesses”, rule 704 “Opinion on an Ultimate Issue”, rule 705 “Disclosing the Facts or Data Underlying an Expert’s Opinion” and rule 706 “Court-Appointed Expert Witnesses”; see [9]) are of little relevance for this paper and are listed here only for the sake of completeness. For a detailed analysis of the relevance of these additional rules as well as the Federal Rules of Criminal Procedure rule 16 section G (“Expert Witnesses”; [11]) in the presentation of forensic evidence via expert testimony see [2].

2.2. Daubert challenges to forensic methods

Regarding the second and third point of the list given above in section 2.1 in the analysis of FRE rule 702 (“Type of knowledge considered” and ‘Who is addressed by the expert’) it has to be summarized that if something is declared to be ‘science’ in regard to FRE rule 702 then the criteria for the evaluation of scientific methods introduced in Daubert have to be applied by the judge to make the expert prove this declaration. These criteria and their relevance for this paper are discussed in this section.

In 1923 the court in Frye v. United States, 293 F. 1013 (D.C. Cir. 1923) made a first suggestion how to proceed with the admission of expert testimony based on novel forensic techniques. The court in Frye suggested: “[...] while the courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.” In Frye (or the Frye standard as it is also referred to) the court concluded that the polygraph test that was intended to be used in this case could not be admitted because it lacked the required general acceptance in the corresponding research fields.

The Frye standard was in 1975 replaced by the FRE. In the original version they contained no special rule that, when dealing with ‘scientific’ evidence, novel or otherwise, ensured that science-based testimony is reliable and, therefore, admissible. Therefore, all evidence was considered admissible if relevant, provided its use in court was not outweighed by “unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time or needlessly presenting cumulative evidence”, as stated in FRE rule 402 [9].

The next relevant step in legal developments on expert testimony (and therefore the means of introducing forensic sciences into court) occurred in 1993, when the U.S. Supreme Court made another ground-breaking decision on expert testimony in Daubert. In 1999 this was followed by another important court case, Kumho Tire Co. v. Carmichael, 119 S.Ct. 1167 (1999) – short ‘Kumho Tire’. Both Daubert and Kumho Tire arose out of civil lawsuits. An extensive and intelligible summary of the proceedings in the Daubert cases (original and the affirmation in the U.S. Court of Appeals) is presented in [2]. The main point of interest for this paper is that the court unanimously held that Frye did not survive the enactment of the FRE. In interpreting FRE rule 702, the court in Daubert stated that if the admissibility of scientific evidence is challenged, it is the function of the trial court to act as ‘gatekeeper’ to determine whether proffered opinion evidence is relevant and reliable. The U.S. Supreme Court specified several flexible and non-exclusive criteria (the so called Daubert criteria or Daubert standard) to guide other courts when they have to consider in deciding whether a scientific field is sufficiently reliable to warrant admission of opinion evidence. As a further important milestone, in 1999 in Kumho Tire the U.S. Supreme Court applied the Daubert criteria of proof of reliability to all forms of expert opinion testimony (i.e. scientific, applied science, technological, skill and experience). Also, the court in Kumho Tire made it clear that the list of Daubert criteria was meant to be helpful and is not a definitive check-list but rather a flexible, non-exclusive recommendation. As a result no attempt has been made in US law to ‘codify’ these specific criteria. Other U.S. law cases have established that not all of the specific Daubert criteria can apply to every type of expert testimony.

The outcome of Daubert and Kumho Tire led in April 2000 to the amendment of FRE rule 702 described in section 2.1. The SWGFAST draws in [2] the following conclusions for forensic scientists from Daubert and Kumho Tire: “[...] The courts may, and many will, require the experts to show that they know what the scientific method consists of and provide the scientific basis for their conclusions. [...] each discipline will be judged by its own standards and upon its own experience. [...] The court can re-litigate the admissibility of a certain type of expert evidence if a litigant can make a credible argument that there has been no previous scientific inquiry of the validity of the assumptions on which a forensic field has long rested. Decades of judicial precedent no longer preclude reviewing whether existing precedent satisfies Daubert and Kumho Tire.”

This summarizing statement, from a forensic expert group trained (and providing training) for court appearances as expert witnesses, defines the requirement for the investigation of the fitness of forensic methods. The actual evaluation criteria are the criteria FRECO to FRECS specified in section 2.1 above as well as the criteria introduced by Daubert for the so-called Daubert hearings.

The specific criteria, explicated by the Daubert court and used within this paper as evaluation criteria for forensic investigations, are [10]:

• DC1: “whether the expert’s technique or theory can be or has been tested – that is, whether the expert’s theory can be challenged in some objective sense, or whether it is instead simply a subjective, conclusory approach that cannot reasonably be assessed for reliability”
• DC2: “whether the technique or theory has been subject to peer review and publication”
• DC3: “the known or potential rate of error of the technique or theory when applied”
• DC4: “the existence and maintenance of standards and controls”
• DC5: “whether the technique or theory has been generally accepted in the scientific community”

While the criteria DC2 to DC5 are self-explanatory (including the fact that publication in DC2 means “open publication”), DC1 is summarized more precisely in as “the theory or technique (method) must be empirically testable, falsifiable and refutable”.

In [2] an extensive and intelligible review of Daubert hearings regarding challenges to the admissibility of friction ridge individualizations (“fingerprints”) for forensic identification of human beings is presented. This review is addressing all Daubert criteria for this field as well as further concerns raised by judges in Daubert hearings. The combination of the criteria directly derived from the FRE rule 702 (FREC0 to FREC3; see section 2.1) and the criteria explicated by the Daubert court (DC1 to DC5) is referred to within this paper as the Daubert criteria. They can be used as performance indicators for an estimation of the currently achieved forensic compliance of a method.

3. DISCUSSION ON THE MATURITY OF A SELECTED APPLICATION SCENARIO

The Daubert criteria are widely accepted in the classical field of the medical forensics (see e.g. [12] and [13]). In digital camera forensics (as one of the most mature research fields in media forensics) a major breakthrough regarding forensic conformity can be seen in the law case United States of America v. Nathan Allen Railey (United States District Court for the Southern District of Alabama, August 2nd, 2011). In the Daubert hearings of this case, the method of photo response non-uniformity (PRNU) based digital camera authentication based on intrinsic characteristics of its image acquisition sensor developed by Jessica Fridrich and her group (see e.g. [14]) got accepted for the first time as forensic evidence. An expert working for the FBIs Forensic Audio, Video, and Image Analysis Unit (FAVIAU) established in the Daubert hearings that this approach meets all necessary criteria (see DC1 to DC5 in section 2.2 above) and the presiding judge furthermore decided that this evidence (or more precisely the FBIs expert testimony based on this media forensic analysis) also meets the FRE rule 702 criteria (FREC0 to FREC3). This is an important success for the whole research field of media forensics. An important fact to mention here that the forensic tool used was not the research prototype provided on the researchers own website (see [15]). Instead the FBI and the Air Force Research Laboratory (AFRL) funded the research and development work necessary to improve the software and incorporate it in a software named “FIND Camera” used now by the FBIs FAVIAU.

In addition to the application of “FIND Camera”, the image source authentication results in the case United States of America v. Nathan Allen Railey were verified by an independent computer forensic examiner with a contract with the U.S. Attorney’s Office in Mobile, Alabama. He successfully matched the electronic serial number embedded by the Kodak camera into all pictures under examination against the camera itself. The result, obviously, supported the findings the FBI experts achieved with “FIND Camera”.

Without having access to the court files, it is impossible to know what exactly the findings of the court regarding the applicability of PRNU-based camera forensics have been in this case. Therefore, the following points summarize the authors own layman’s assessment of PRNU-based camera forensics:

• FREC0, FREC1 and FREC3: Criteria which cannot be answered in general, because they are related to the specific court case under consideration.
• FREC2: The methodology is based on intrinsic characteristics of the used image sensory and has been published upon many times by different research groups. The evaluation methodology for the performance of the software is explained in detail in publications such as [14].
• DC1: In the context of the case United States of America v. Nathan Allen Railey, an expert from the FBIs FAVIAU testified that the government tested the program it developed (i.e. “FIND Camera”) using the evaluation method described in [14] with a million pictures for thephoto-sharing website Flickr using the corresponding Exchangeable Image File Format (Exif) meta-data for plausibility verification. He further testified that the result of these evaluations determined the false-positive rate to be less than one in a million. Additionally, a small scale test on the intra-class variance was made by him using an identical Kodak camera.
• DC2: The methodology has been published upon with a publication count larger than 20 with publications such as [16], [17] and [18] remarks on DC1 above.
• DC4: This point is hard to answer for the author. Obviously, there exist internal procedures at the FAVIAU, also regarding the application of “FIND Camera”, but no corresponding documents are freely found.
• DC5: See remarks on DC2 above.

4. CONCLUSION

The assessment made by the authors in section 3 on an exemplary selected media forensic approach is obviously far away from being complete or sufficient to pass any Daubert hearing. A complete discussion on one hand would by far exceed the page limitations of this paper and on the other hand such undertaking is far beyond the legal expertise of the author. For a clearer picture on how an entire Daubert hearing looks like, the author refers to [2], where one such hearing on the analysis of latent fingerprint traces is discussed by the experts of the SWGFAST. Nevertheless, with the two short assessments presented above the following points can be highlighted:

• Media forensics are strongly depending on the specific data format and in many cases also content-sensitive. This makes

---

3 For a summary of the relevant part of the proceedings see: http://blog.al.com/live/2011/07/expert_witnesses_link_camera_t.html
4 This media forensics approach is based on the fact that photo response non-uniformity (PRNU) noise is usually the dominant part of the pattern noise of a digital camera. It is caused mainly by pixel non-uniformity, a signal primarily concentrated in high spatial frequencies. Light refraction on dust particles, on optical surfaces, and the CCD element itself are low frequency signals that also contribute to the PRNU noise and tend to increase this individualizing characteristics of a digital camera.
it in the context of FREC0 important to clearly specify for a media forensics approach to which type of media and which type of content it can be applied. These specifications have to be communicated clearly to the researchers to the corresponding forensic investigators.

• The actual intention of DC1 (in combination with FREC1 and FREC2) is to help judges to keep ‘junk science’ out of their courtrooms (see the discussion on the non-admissibility of polygraph tests summarized in [2]). Despite the fact that it can be assumed that any researcher has the intrinsic interest to test his work, DC1 implies for us that very high standards should be set for the evaluation of new principles and methods as well as for the documentation of the evaluation results and the evaluation methodology.

• Scientific results on media forensic techniques must be published and openly discussed with the corresponding community otherwise no compliance with DC2 and DC5 can ever be achieved. DC2 is in strong accordance with the general principles of academic research, where ‘publish or perish’ rules entire communities. Industrial research, on the other hand, still often enough prefers ‘security by obscurity’ as main driver for its publication strategy, which is counterproductive if it comes to intended admissibility in court. Furthermore, scientific results on media forensic techniques should be communicated to a wider audience – preferably the general public – to counter the ‘CSI effect’ in courts.

• Regarding the error rate of techniques, DC3 aims to enforce statistically relevant evaluation setups. For a researcher an evaluation with 20 or 30 test subjects is sometimes already enough to propose a new method for discussion in the corresponding community (usually by paper submission to a conference or workshop) but for court admissibility it is obvious that higher standards of maturity have to be met in the evaluation or benchmarking process.

• For many academic researchers the idea of investing time and effort into standardization work is a nightmare. Nevertheless, DC4 (in combination with FREC3) strongly motivates us to do so.

• The main point of DC5 (whether the proposed technique or theory is generally accepted in the corresponding scientific community) is hard to assess for a researcher. Here, we have to help practitioners like forensic investigators to identify other researchers working on similar topics who can establish this fact. Here we have to acknowledge the fact that if a media forensic technique sees field application (e.g. by application in court trials), it will sooner or later have to deal with countermeasures – so called attacks or anti-forensics.

For the PRNU example this is very well illustrated in [19] where a PRNU fingerprint from a different camera is ‘transplanted’ to a digital image. Therefore, the compliance to DC5 has to be re-evaluated on a regular basis.

Summarizing the work presented here, it has to be emphasized by the author that media forensics – even though it receives currently enormous research interest – still has to be considered to be very immature. Initial work, like the one discussed here or in [20], needs to be enhanced to allow for full exploitation in court.

5. REFERENCES