Forensic Science International: Digital Investigation 32 (2020) 300918

Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

DFRWS 2020 EU – Proceedings of the Seventh Annual DFRWS Europe

IP Addresses in the Context of Digital Evidence in the Criminal and Civil Case Law of the Slovak Republic

Pavol Sokol [a, *], Laura Rózenfeldová [b], Katarína Lučivjanská [a], Jakub Harašta [c]

[a] Pavol Jozef Safarik University in Košice, Faculty of Science, Jesenná 5, Košice, Slovakia
[b] Pavol Jozef Safarik University in Košice, Faculty of Law, Kováčska 26, Košice, Slovakia
[c] Masaryk University, Faculty of Law, Veveří 158/70, Brno, Czech Republic

ARTICLE INFO

Article history:

Keywords:
IP address
Digital evidence
Criminal and civil proceedings
Privacy
Personal data
Anonymisation

ABSTRACT

Use of IP addresses by courts in their decisions is one of the issues with growing importance. This applies especially at the time of the increased use of the internet as a mean to violate legal provisions of both civil and criminal law. This paper focuses predominantly on two issues: (1) the use of IP addresses as digital evidence in criminal and civil proceedings and possible mistakes in courts' approach to this specific evidence, and (2) the anonymisation of IP addresses in cases when IP addresses are to be considered as personal data. This paper analyses the relevant judicial decisions of the Slovak Republic spanning the time period from 2008 to 2019, in which the relevant courts used the IP address as evidence. On this basis, the authors formulate their conclusions on the current state and developing trends in the use of digital evidence in judicial proceedings. The authors demonstrate the common errors that occur in the courts' decisions as regards the use of IP addresses as evidence in the cases of the IP addresses anonymisation, usage of the in dubio pro reo principle in criminal proceedings, and the relationship between IP addresses and devices and persons.

1. Introduction

Digital forensics aims to answer the 'what', 'why', 'how', 'who', 'where' and 'when' type of questions (Montasari, 2017). In some cases, these answers are later presented to the court. This process engages many stakeholders with different interests, different level of knowledge and different roles (Stahl et al., 2012). As demonstrated by Brungs and Jamieson (2005) and Liles et al. (2009), members of these groups often disagree on the importance of specific legal issues related to digital forensics. These results extrapolate to the area of general issues as well, as digital forensics experts, members of law enforcement agencies, attorneys and judges focus on different issues and face different challenges.

Additionally, these stakeholders' groups have a different level of influence over existing forensics practices. Forensic experts are required to satisfy legal requirements. This led to a surge in development of tools aiming to ease the communication of basic concepts to the audience with legal background. On the side of forensic experts, standard operating procedures are developed, such as the procedure for router examination at the scene presented by Horsman et al. (2019). Standard operating procedures and similar tools ensure '*the validity, legitimacy and reliability of digital evidence*' (Slay et al., 2009) as these are the core values behind any forensic process (Vincze, 2016). Development of those procedures leads to standardisation across the investigation to ensure '*that work is done consistently by all persons who are required to do the same task*' (Manghani, 2011). The ability to seize digital evidence and to do so legally is often followed by the challenge of making lawyers understand the evidence. Lawyers and judges must be able to grasp the basic concepts in order to use the information provided to them by forensic experts efficiently. Education is often discussed, as a mean to allow lawyers to understand what they can expect from digital forensic experts and even what questions should lawyers ask them (Wong, 2013) (Oparnica, 2016) (Henseler and van Loenhout, 2018). In the past, limited education of lawyers was quoted as one of the reasons for underuse of digital evidence in the USA (Rogers et al., 2007).

The issue of 'how' to conduct the digital forensics investigation is shaped by the legal requirements. These requirements may be difficult to pinpoint, as they are shaped by legal acts and by the practice of courts. These are understandably prone to change and development. Analyzing court decisions can lead to the

* Corresponding author.
E-mail address: pavol.sokol@upjs.sk (P. Sokol).

https://doi.org/10.1016/j.fsidi.2020.300918

identification of legal requirements for seizing and handling digital evidence. Additionally, in the case of the absence of a unified courts' practice, analysis can lead to the identification of gaps in the use of digital evidence, which can, in turn, lead to legislative action clarifying the requirements. Such analyses depend largely on the availability of courts' decisions for further analysis. In our research, we conduct a quantitative analysis of the use of IP addresses as a special type of digital evidence by Slovak courts. Our goal is to find out how IP addresses are represented as evidence in court decisions and how various issues (identifiability, multiple users using the same device, the difference between IPv4 and IPv6 etc.) are represented in these decisions.

The main goal of this paper is to analyse different parameters that impact the courts' final decision when using the IP address as evidence. To achieve this goal, we aim to answer the two research questions.

1. What attributes are important for court when it needs to rely on IP address as evidence? What are attributes important to the court's decision? In particular, what evidence with respect to IP addresses is relevant?
2. How courts approach the anonymisation of IP addresses in their activities?"

The second aim of this paper is to consider the courts approach to the IP addresses' anonymisation, specifically when considering IP address as an online identifier. In this regard, we refer to the relevant judgements of the CJEU, especially to the case C-582/14 Breyer and the subjective approach it formulated to answer the question of whether IP addresses are able to identify a natural person.

This paper is organized into five sections. Section 2 focuses on the review of the published research related to the research questions. Section 3 provides details on the research methodology and outlines the data set and methods used for the analysis. Section 4 describes the results of the analysis and discusses key takeaways obtained from the analysis. The last section contains conclusions and our suggestions for future research.

2. Related work

Related work contained in this part is formative for our study from three different angles. First, our paper complements explanatory studies focused mainly on a qualitative analysis of available legal and policy documents that form states' response to the rise of cybercrime and the rise in the use of digital evidence in general. Second, our work is related to sources explaining the legal nature of IP addresses. Finally, our paper is also relevant with respect to studies engaging in qualitative or quantitative research into the use of digital evidence by courts in different countries.

2.1. Exploratory analysis of legal documents

The first international treaty seeking to harmonize national laws, to improve investigation techniques in individual states and to improve cooperation between these states when investigating and prosecuting cybercrime is the Budapest Convention on Cybercrime. It is important to note that a lot of variants for national laws, even as framed by this convention, still exist. Overall, there are many studies which focus on explaining the specifics and idiosyncrasies of legal requirements in different countries. Studies focusing on individual countries often approach the issue in their complexity – such is the case of Shukan et al. (2019) who focused on the overall issue of cybercrime control in Turkey—or focus on individual issues providing wider audience with explanation of national laws for further comparative studies – such is the case of Abu Issa et al. (2019) focusing on the specific crime of unauthorized access and its prosecution in Jordan.

Similar studies have a limited impact. Complex comparative studies are quite rare and often prepared by experts under auspices of international organisations to compare legal provisions in their member states, e.g., Council of Europe's European Committee on Legal Co-operation (On Legal Co-Operation, 2016). However, the most attention of international scholars and their readership is often dedicated to legal requirements in the USA and the United Kingdom. Montasari (2017) identified key documents and court decisions related to both the United Kingdom and the USA in the area of disclosure and admissibility of evidence. Further non-exhaustive list of studies focusing on legal requirements in these jurisdictions includes oft-cited papers of Ryan and Shpantzer (2002), Wegman (2005), Nance and Ryan (2011), Garrie (2014) or Cole et al. (2015). These studies mostly identify key court decisions shaping the practice and forming the legal requirements for presenting evidence before (primarily) US and (secondarily) UK courts. Strong focus on US courts and courts in the United Kingdom might be driven by the availability of court-decisions in these countries, language issues and strong position of court decisions within the precedential system. Precedents often tackle relatively specific issues compared to the existing legal acts typical for non-precedential countries.

Based on this part of related work, we conclude that studies focusing on less significant countries must go beyond the mere description of legal requirements or small-scale comparative notes. Preferably towards qualitative and quantitative studies of how digital evidence or any specific type of digital evidence is handled by courts. Understanding the court practice on a larger scale could yield interesting insights even in countries that lack strong and formative case law with precedential force in the area of digital investigation.

2.2. Legal nature of IP address

The question of whether IP addresses are to be considered personal data, and if so, under what circumstances, is a highly debated topic among the concerned parties (internet service providers, website operators, legal professionals and scholars).

Lundevall-Unger and Travik (Lundevall-Unger and Tranvik, 2010) examined the nature of IP addresses and discussed the tools and methods to be used to analyse the applicability of the identifiability criterion. They also proposed a practical method for deciding the legal status of IP addresses with regard to the concept of personal data based on the legality test and the likely reasonable test (Lundevall-Unger and Tranvik, 2010). Schwartz and Solove (1814) discussed mainly the capability of a natural person's identification based on IP address, emphasizing that such identification is naturally indirect. They examined the issues connected with the process of identifiability with special attention to the possibility of multiple users accessing the same device. This issue was later re-examined by Stalla-Bourdillon et al. (2016). In this study, the authors concluded the need for additional data to establish the IP address as identifying a specific person using the device.

Weber and Heinrich (2012) considered the traceability of IP addresses, information represented by IP addresses and the possibility of natural persons' identification in relation to both static and dynamic addresses. Banterle (2018) focused on the description of the former approach to the evaluation of the nature of indirect identifiers as dynamic identifiers, based on circumstances of specific cases. Von Grafenstein (von Grafenstein, 2018) discussed the differences between static and dynamic IP addresses. He focused mainly on the context of functional differences between the IPv4 and IPv6 IP addresses. In the case of IPv6 IP addresses, each device

receives a single unique address, which adds further layers of complexity in the relation between a natural person and IPv6 IP addresses. Thus, he challenged the dominant notion of IP address being personal data under existing data protection legislation. Similar path, but with the accent to different issues were followed by Tamó-Larrieux (2018), who illustrated the possibility to avoid identification based on IP addresses with the use of a proxy server.

Article 29 Data Protection Working Party in its Opinion 4/2007 on the concept of personal data (WP29, 2007) adopted a position, that although IP addresses may in some cases be considered as personal data while in others not, there is an imminent risk involved that internet service providers will not be able to distinguish these cases in real-time. The opinion stated that in the absence of absolute certainty on whether the IP address can be used to identify a specific user, it would be required to treat all IP addresses as personal data '*to be on the safe side*'. It shed additional light on this issue in the judgment of the Court of Justice of the European Union (hereinafter as 'CJEU') in the case C-70/10 Scarlet Extended. In this case, the CJEU stated in paragraph 51 that IP addresses are *'protected personal data because they allow (…) users to be precisely identified'* (CJEU-70/10, 2011). In this case, the IP addresses in question were of static nature and were, together with additional data necessary to identify a specific person, in possession of a single subject, an internet service provider. In its later judgment in the case C-582/14 Breyer, the CJEU approached the issue of dynamic IP addresses and, following the recital 26 of Directive 95/46/EC, formulated a test to determine whether the dynamic IP address is to be considered as personal data. CJEU considered in paragraph 45 (similarly to (Lundevall-Unger and Tranvik, 2010)) *'whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constituted a means likely reasonably to be used to identify the data subject'* (CJEU-582/14, 2016). In this context, the CJEU formulated a subjective approach in determining the personal data nature of IP addresses deriving it from the opportunities and possibilities of a subject to identify a natural person.

As is evident from related work and the existing cases in this area, the concept of IP address in terms of its ability to identify a single individual came through different phases. These phases were shaped by the evolution of the CJEU's practice, as well as by technological advancement (IPv4 vs IPv6, rise in use of proxy servers).

2.3. Use of digital evidence

Significant attention has been recently drawn to the questions of how the legal framework interacts with technological development and how this interaction forms the general practice for digital investigation. Legal requirements – in Pound's term law in books – interact with other phenomena shaping the judges' decision-making process to form the law in action (Pound, 1910).

A research report by Goodison et al. (2015) summarized the Workshop on Digital Evidence Needs, which was held in 2014 to identify weak points in digital forensics that require further attention of relevant stakeholders. Some of the identified issues related to legal requirements and interaction between lawyers and digital forensic experts [8, p.18–21].

James and Gladyshev (2016) surveyed members of the law enforcement agencies through an online survey. They provided the qualitative and quantitative analysis of the answers to understand how the mutual legal assistance works, how well it is known, how often it is used, through what channels and to what results. Authors noted a lack of best practice as one of the factors leading to inconsistent results in both sending and receiving mutual legal assistance requests.

Gřivna and Drápal (Gřivna, Drápal) analysed court decisions of the Czech courts to understand what types of cybercrime are prosecuted and to what ends. In their conclusion, they noted that criminal offences associated with cybercrime that are dealt with by the Czech routs are not representative of the breadth of cybercrime. This leads the authors to conclude that the whole system could be more effective in prosecuting complex criminal offences, as well as to suggest certain changes in the existing legal definitions of cybercrime.

The related work in this area suggests that the law in action aspect of dealing with digital evidence is of the same importance as is the law in books. The static legal requirements formed by legal acts often come together with other factors. In countries lacking strong precedential decisions, such as Slovakia, analysing larger datasets of court decisions with the focus on their qualitative and quantitative aspects could lead to interesting insights into the effectiveness of obtaining and handling digital evidence.

3. Methodology

The analysis required the collection of all of the judicial decisions of the courts of the Slovak Republic (courts of the first instance, of the second instance, the Supreme Court and the Constitutional Court) spanning the time period from 2008 to 2019, in which the term 'IP address' was present (in total 398 decisions). These decisions were collected from the automated system of legal information - ASPI (Kluwer, 2019), which contains more than 3 million decisions adopted by the courts of the Slovak Republic. The analysis of these 398 decisions has led the authors to exclude 211 decisions from the final analysis. These included predominantly decisions, in which the term IP address was used only to describe e. g. a service provided by the operator (dynamic IP addresses provided as a part of the overall service of internet connection), where the recipient of such service failed to pay the invoice. Decisions from both civil and criminal proceedings were collected. However, decisions from certain types of proceedings were excluded from the analysis (e. g. judgement enforcement proceedings, bankruptcy proceedings). In these decisions, the IP address was only briefly mentioned, without being used as evidence by the court and without the exact specification of its numerical form, therefore not considering the issue of IP's addresses anonymisation. The authors, therefore, concluded that these decisions provided no value to the analysis, as it was not possible to determine the individual attributes relevant for the analysis. Once we removed these decisions, we were left with a total of 187 decisions for further analysis.

The analysis included the collection of the metadata for each of the decisions, in particular (I) the docket number, (II) date of proceedings' beginning, (III) date of proceedings' end, (IV) type of proceedings (civil/criminal), (V) name of the deciding court, (VI) type of the deciding court in the court hierarchy, (VII) outcome of the case (guilty/not guilty in criminal cases, successful/unsuccessful action in civil cases), and (VIII) number of occurrences of the term 'IP address'.

In order to collect data to answer the first research question regarding the decision-making process of courts in connection with the use of IP addresses, we have collected the following data from court decisions:

- specification of the IP addresses' timestamps related to specific date and time; attributes: yes - no;
- IP addresses' assignment, where we considered whether the IP address in question was understood as assigned to the person or device; attributes: person - device - person and device - not assigned;
- further specification of a device tied to IP address (brand name and type, serial number, IMEI number etc.); attributes: yes - no;

- consideration of a person's relation to the device in question (e.g. identification of ownership); attributes: yes - no;
- court's consideration of the possibility of other people accessing the device; attributes: yes - no;
- type of the device in question; attributes: home - work - public - other;
- IP address only mentioned without further information; attributes: yes - no;
- evidence considered in the case[1]; attributes: testimonial (defendant/witness), documentary, real, expert opinion.

In order to answer the second research question regarding whether IP addresses are treated as personal data while publishing the court decision, we have collected the following data from court decisions:

- anonymisation of 'traditional' personal data (e.g. surname, date of birth) in the decision; attributes: yes - no;
- anonymisation of other online identifiers (e.g. mail address, login details) in the decision; attributes: yes - no;
- anonymisation of IP address; attributes: yes - no.

On top of these data, we have also distinguished whether the court when deciding the case differentiated between types of IP addresses. Namely:

- static and dynamic IP addresses;
- public and private IP addresses;
- IPv4 and IPv6 IP addresses.

For the purpose of geospatial analysis, we used external geospatial service ip-api.com. All non-anonymised IP addresses were enriched by spatial data (e.g. latitude, longitude, country, internet service provider, time zone) using this service.

For the quantitative analysis, we transform yes/no attributes to binary variables taking values of 1 (yes) or 0 (no). Categorical variables taking several different values were also transformed into binary variables. For example, the variable of IP addresses' assignment to person or device is transformed to two binary variables - *person* with an assigned value of 1 or 0, and *device* with an assigned value of 1 or 0.

At the first stage of the analysis, we look at the standard measure of the dependence between two variables—the correlation coefficient (Lee Rodgers and Nicewander, 1988). The values further away from zero indicate a stronger relationship. Positive values provide evidence that the higher value of one variable is associated with a higher value of the second variable. Negative values indicate that the higher value of one variable is associated with a lower value of the second variable.

To study how different types of evidence contribute to the final decision of the court, we employ the logistic regression (Menard, 2002). Thus, we analyse the significance of individual attributes and their explanatory power to explain the court's decision. We rather prefer the logistic regression to the standard regression. It is more appropriate for our first research question because the variable for which we aim to find explanatory attributes is binary (court's decision).

Finally, we try to identify specific groups of cases. For that, we employ the hierarchical clustering method (Kassambara, 2017) which is especially appropriate for the binary data we use to describe individual cases in our dataset. First, we assess the clustering tendency by the Hopkins statistics and the visual inspection of the two-dimensional projection of the dataset (Fig. 1). Second, we interpret the four groups obtained from the hierarchical clustering method clustered based only on attributes (not decisions).

4. Results and discussion

In this section, we focus on analysing the decisions themselves in the context of the goals set. We will look more closely at the relationship of the individual attributes of the decisions to the outcome of the decision, as well as the relationship between the attributes. We will then discuss privacy and personal data issues in the context of using IP addresses. The last point is a look at the spatial analysis of IP addresses and their use of such an approach.

4.1. Correlation analysis of attributes

The heat map in Fig. 2 reports the correlation coefficient for all pairs of attributes. It shows the relationship (or the lack thereof) between the individual attributes as defined in the methodology. As the most obvious relation, we can distinguish the direct dependency that exists between the uses of particular types of evidence. To illustrate, if an expert opinion was produced and specified in the judgement's reasoning, there is a high probability that other types of evidence were also produced and considered by the court.

Dependency also exists between the use of expert opinion and a closer stipulation of the person's relation to a device. The dependency, with a correlation coefficient value of 0.22, indicates that this stipulation may have firstly been examined in the expert opinion itself.

A high level of correlation, with a value of 0.56, also exists between the assignment of an IP address to a device and the device's specification. If the court correctly assigns IP addresses as device identification and not person identification, it is more likely that the device in question will be identified and further specified in the decision.

Moreover, we observe a high level of correlation, with a value of 0.55, between the device's specification and the relationship between the device and a person. As regards the assignment of an IP address to a device, our analysis has shown that such assignment (whether to a person or a device) have not occurred in all of the decisions examined. In the context of cases in which such assignment can be found (in total 89 decisions), in 86,5% of them (77 cases) the IP address was correctly assigned to a device, while only in 9 of them directly to a person. What is interesting is the fact that we were also able to identify three decisions in our dataset, in

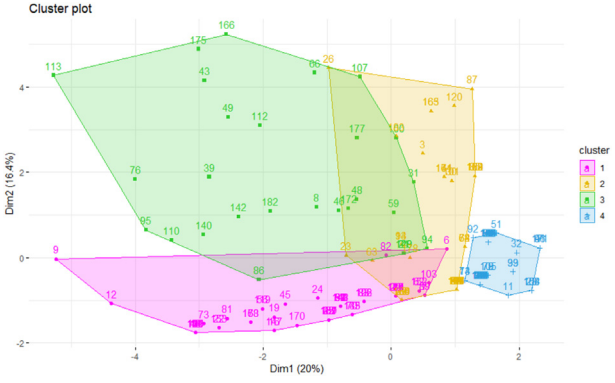

Fig. 1. Projection of the cluster to 2-dimension space.

[1] Evidence types are stipulated in the Slovak Act No. 301/2005 Coll. on Criminal Procedure. These include the testimony of a defendant, testimony of a witness, documentary evidence (written statements, reports and other documents), real evidence (including audio and video recordings), and expert opinions.

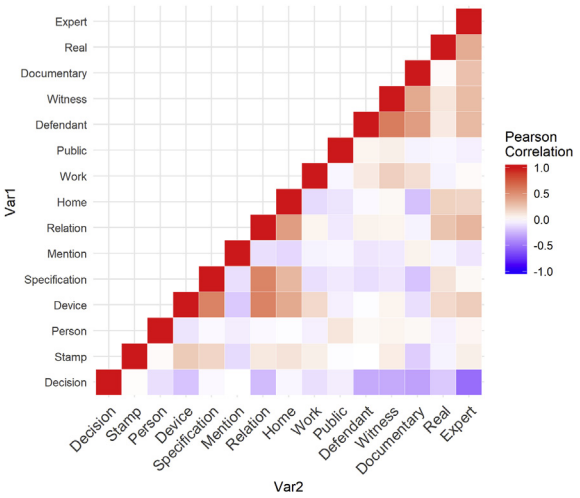

Fig. 2. Heatmap - correlation analysis of attributes. Attribute 'Expert' means expert opinion, attribute 'Real' means real evidence, attribute 'Documentary' means documentary evidence, attribute 'Witness' means testimony of a witness, attribute 'Defendant' means testimony of a defendant, attribute 'Public' means device with IP address was in public, attribute 'Work' means device with IP address was at work, attribute 'Home' means device with IP address was at home, attribute 'Relation' means person's relation to the device, attribute 'Mention' means IP address only mentioned without further information, attribute 'Specification' means further specification of a device tied to IP address, attribute 'Device' means address was assigned to the device, attribute 'Person' means IP address was assigned to the person, attribute 'Stamp' means specification of the IP addresses' timestamp, attribute 'Decision' means final decision (guilty or successful action).

which the IP address was assigned both to a device as well as to a person. One of these cases was especially interesting, as the IP addresses' assignment was specified directly in the decision's verdict, in which two offences were contained, in one of which the IP address was assigned to a person, while in the other to a device. As this decision was in the form of a criminal order and therefore lacked reasoning, we were not able to identify the intention behind such distinction. However, this correlation between the device's specification and the establishment of a relationship between a certain person (usually a defendant) and such device can suggest the court's inclination to firstly determine the device, to which the IP address was assigned, and only later to consider who had access to the device in question (whether it was the defendant herself or whether other people were also able to access the device). This factor will be of utmost relevance in the criminal proceedings, where the court must establish the defendant's guilt, considering the in *dubio pro reo* principle ([when] in doubt, for the accused).

A high level of correlation, with a value of 0.18, can be found between the device's specification and the stipulation of the IP addresses' time stamp. In this regard we have expected that in each case the IP addresses' timestamp should be specified. However, this was not confirmed by our analysis. In only 54% of the cases (101 decisions), both the time and date of the IP address' use were enumerated. In 85 decisions, no such enumeration was provided. In one case, the court specified in its decision time stamps in connection to some, but not all IP addresses.

Fig. 2 also demonstrates interesting dependencies between the decisions themselves and certain attributes. To illustrate, expert opinions used in the court's reasoning usually lead to the judgement of acquittal in criminal proceedings or to the judgment dismissing the action in civil proceedings. Different types of evidence, as well as a closer specification of the person's relation to the device and the IP addresses' assignment to the device, have similar effects, although on a smaller scale.

Coming back to the in *dubio pro reo* principle, according to which the court is required to construe factual circumstances of the case in favour of the accused in case of doubts, the null hypothesis (presumption) is that the court is expected to substantiate its deliberation on guilt/action on different evidence. Therefore, it is possible to assume the existence of a direct dependency between conviction in criminal proceedings or a successful action in civil proceedings and the evidence produced.

To study this aspect empirically, we constructed a new attribute - counting the number of types of evidence considered (testimonial, documentary, real, expert opinion). Thus, the minimum value is zero, and the maximum value is 4. The correlation coefficient of this variable with the attribute *guilty* is −0.49, indicating a strong negative relationship - more pieces of evidence are followed by the decision of no guilt. We support the evidence by running the logistic regression with the attribute and find that the number of evidence has a significantly negative effect (with the p-value less than 0.001). In sum, both types of the analysis suggest that the more different types of evidence are used, the more likely it is that a person will be found not guilty or the civil action will be unsuccessful.

In our research, we have focused on the possibility of creating the examined decisions' profiles based on the different attributes chosen as specified in the methodology. In this part of our analysis, we have only considered those decisions, in which the court provided the reasoning for the verdict adopted. The reason for this is the fact that numerous decisions did not contain the reasoning part at all, as § 163 (1) (d) of the Slovak Act No. 301/2005 Coll. on Criminal Procedure allows with regard to certain types of decisions as stipulated by law. This is, for example, the case of decisions in the form of an agreement on crime and punishment, where the defendant admits her guilt, and the court further examines only the evidence necessary to stipulate the scope and type of punishment and not the evidence on the crime and question of guilt. Another type of a decision, for which no reasoning is provided, is the so-called criminal order. In this case § 353 (1) of the Act No. 301/2005 Coll. on Criminal Procedure provides the court with the opportunity to issue a decision without hearing a case in the trial if the body of crime is sufficiently established by investigation. Criminal order is a judgement of conviction, and the defendant is provided with a specific type of a remedial measure - protest. It must be noted, that from all of the decisions examined, 42 were in the form of an agreement on crime and punishment, another 42 decisions were in the form of a criminal order and in 26 of the decisions the defendant and the prosecutor waived their right to appeal the court's decision which triggers the application of § 172 (2) of the Act No. 301/2005 Coll. on Criminal Procedure, according to which in such situations the court only formulates a simplified version of the judgement (without reasoning).

The decisions, in which the reasoning part was present, were analysed with the use of the clustering method. The results led to the differentiation of the decisions examined into the following 4 clusters (Table 1).

The first cluster of decisions is characterized by the high probability of including the IP addresses' timestamps in the decisions (71%). In this group of decisions, the IP address is assigned to a device (with the probability of 95%), which is further specified in the decision and is usually found in the defendant's home. The closer specification of the defendant's relationship to the device is also more probable within these decisions (39%). As regards the evidence, the court is less likely to consider other evidence in the proceedings (4% real evidence, 2% expert opinion).

In the second cluster of decisions, the timestamps are specified in almost all of the decisions (87%). However, no assignment to a device (and therefore no specification of such device) is provided.

Table 1
Table of clusters.

Cluster	Cl.1	Cl.2	Cl.3	Cl.4
Number of decisions	56	47	28	56
Final decision (guilty or successful action)	98%	89%	46%	96%
Timestamp	71%	87%	54%	9%
Assignment to person	4%	4%	4%	13%
Assignment to device	95%	0%	96%	0%
Spec. of device	54%	0%	21%	0%
Only mentioned	0%	0%	0%	14%
Relationship	39%	0%	50%	0%
Location at home	55%	21%	32%	0%
Location at work	4%	0%	18%	0%
Location in public	2%	2%	0%	4%
Testimony of a defendant	0%	30%	36%	0%
Testimony of a witness	0%	19%	39%	0%
Documentary evidence	0%	43%	93%	41%
Real evidence	4%	2%	14%	0%
Expert opinions	2%	11%	57%	0%

In 4% of cases, the IP address used is assigned to a person. Moreover, it is likely that the IP address is localized to the defendant's home (21%), or in a limited number of cases to public places (2%). For these decisions, a relatively high likelihood exists with regard to the use of different types of evidence, such as documentary evidence (43%), defendant's testimony (30%), witnesses' testimony (19%) and expert opinions (11%).

As regards the third cluster of decisions, here we can distinguish decisions in which the IP address used is in almost all of the cases assigned to a device (96%), the relationship between the defendant and the device is usually considered (50%), while the specification of the device is less likely (21%). However, more than half of these cases include the specification of the IP addresses' timestamps (54%). Moreover, the IP address is usually localized in the defendant's home (36%) or at her place of work (18%). This cluster is also characterised by the high probability of the use of a combination of different types of evidence in the decision's reasoning, specifically documentary evidence (93%), expert opinions (57%), witnesses' testimony (39%), defendant's testimony (36%) as well as real evidence (14%).

In the last cluster of decisions, the IP address can be used in the decision's reasoning, only marginally (14%), without a closer examination of different attributes relevant to it. Here we can determine the highest probability of assigning the IP address to a person (13%), without providing further information, such as timestamps (only in 9% of cases). Moreover, in this regard, the court usually only additionally considers documentary evidence (41%).

If an IP address is assigned to a device, a higher probability exists that the relationship between the device and the person will be specified. This confirms our view that the court is trying to connect the online activity by connecting IP addresses with the devices to which they are assigned and only later to connect these devices to a specific person.

4.2. Privacy and personal data protection issues

The applicable legislation on personal data protection is currently contained in the General Data Protection Regulation (hereinafter only as 'GDPR'). GDPR determines its applicability by answering a simple question: what information constitutes personal data? This simple question, however, cannot be easily answered, as certain information may, in some cases, be considered as personal data, while in others not. Whether certain information is to be considered as personal data is determined by applying the so-called 'identifiability criterion', which requires the individual assessment of every information and whether it allows for the identification of a natural person. Moreover, it is necessary to distinguish between information allowing direct or only indirect identification of a person. Similarly, we must consider the subjective criterion formulated by the CJEU in its judgement in the case C-582/14 Breyer, according to which the nature of the subject in possession of information examined (and other data at its disposal) is relevant in determining whether such information will be regarded as personal data or not.

In our analysis, we examined how the courts of Slovakia approached the issue of personal data anonymisation in their judgements, in which the term 'IP address' was contained. We distinguished personal data included in these decisions into three categories: 'standard' personal data, other online identifiers and IP addresses.

As regards 'standard' personal data, these present in our analysis information relating to an identified natural person, or in other words, information allowing for direct identification of an individual without the need to possess further data. These include information such as a surname, date of birth, birth identification number, personal ID number etc. Anonymisation of 'standard' personal data should not pose a problem for the courts, as the nature of this information as personal data is generally accepted. This presumption was confirmed by our analysis, as in almost all of the cases (in 181 decisions) were these 'standard' personal data anonymised. In a number of cases (6 decisions), however, we were able to find the name of the defendant. The reason for this is, in our opinion, that the subject responsible for the anonymisation process made a mistake and forgot to anonymise this data, as this only happened in the case of multiple pages long decisions and in the reasoning part of the decision (and not in the decision's beginning, where the defendant's name is stated, which was anonymised in all of these decisions).

In connection to online identifiers, GDPR recognizes the possibility to identify a person indirectly, for instance, by reference to different identifiers, such as online identifiers. Recital 30 of GDPR expressly acknowledges the possibility to use traces left by different online identifiers provided by the persons' devices, applications, tools and protocols in combination with additional data (unique identifiers or other information received by the servers) to create profiles of natural persons and identify them. Interestingly, this explanation of online identifiers was adopted in the form of a legal definition in § 5 (k) of the Slovak Act No. 18/2018 Coll. on personal data protection.

In our analysis, we have also gathered data on what types of online identifiers were used in the decisions examined (other than IP addresses), as well as whether such identifiers were anonymised or not. Our analysis showed that the most commonly used online identifier (included in 69 decisions) was login information (e.g. account names for social media websites, discussion forums or other pseudonyms used to identify a specific user on a certain website and allow him/her to log in). The second most used online identifier was an email address (present in 26 decisions). Other online identifiers included information such as an ICQ number (1 decision) or an ID number assigned to a customer (1 decision) etc.

As regards the anonymisation of other online identifiers, it is important to note that not all of the decisions analysed contained an online identifier different from an IP address. Other online identifiers could be found in 112 decisions, the majority of which were anonymised (83 decisions, which account to 74% of the decisions examined). In 25 decisions, the anonymisation process did not occur at all. In 5 cases, which contained more than one online identifier, the court decided to anonymise some online identifiers, while not the others. In this context, usually an email address was anonymised, while a user account name was not. The reason for this is, in our opinion, that the user account names in these

decisions were not considered by the court as allowing identification of a natural person, as they only included random numbers or letters or different names and surnames commonly used in Slovakia. In one case, however, the account name of the defendant was not anonymised, while the account name of the person, with whom the defendant communicated, was anonymised. Unfortunately, as we do not possess the decisions in their original form (before the anonymisation process occurred), we are not able to provide further analysis of the courts' decision-making process when it comes to anonymising other online identifiers.

The decisions analysed differ significantly when it comes to the anonymisation of the numeric representation of IP addresses. From all of the decisions examined (187 in total), the analysis showed that in almost 61% of them, the IP addresses present were anonymised (113 decision), a little over 30% of the decisions were not anonymised (57 decisions) and 5% of the decisions did not specify the numerical representation of the IP address (10 decisions). In the remaining 3% of the decisions examined (6 in total), in which more than one IP address was present, some IP addresses were anonymised, while others were not. We were not able to find any specific reason for the court's decision to anonymise some, but not all of the IP addresses present. In four of these decisions, which were the result of a criminal proceedings and dealt with more than one criminal offence, IP addresses were anonymised in some, but not in all of the offences as defined in the court's verdict, or they were anonymised in the courts verdict, but not in the decision's reasoning. Similarly in the civil cases, in which the IP address was used to identify a contractual party to a contract concluded online, some IP addresses were anonymised, but others were not, without any distinction made between them by the court in the reasoning. Other than a possible mistake made during the anonymisation process, we were not able to find any common factor which would justify such an approach.

Fig. 3 illustrates the fact that IP addresses are not anonymised, which is a discrepancy between how IP addresses are viewed by 'law in books' (IP address as personal data) and how the 'law in action' works with them (necessity to anonymise personal data in certain cases). This demonstrates that the application of law in practice is often not what the applicable law foresees. The adoption of GDPR (April 2016) and the CJEU's case C-582/14 Breyer (October 2016) provided further insights as regards the conditions under which IP addresses are to be considered as personal data. As we can see in Fig. 3, the way IP addresses are viewed is changing, however different views may be distinguished in the context of law in books and law in action.

4.3. Discussion of the related issues

A part of the forensic analysis focused on issues concerning legal geography (Where), specifically considering the question of the country's jurisdiction (Ieong, 2006). As regards IP addresses, we examined the localization of the IP addresses that were not anonymised (from all the decisions examined, in 126 decisions, no anonymisation occurred). Our conclusions are based on this analysis.

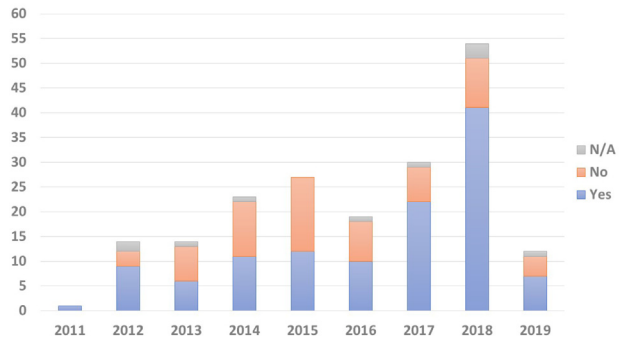

Fig. 3. Anonymisation of the IP addresses in the decisions.

The majority of the non-anonymised IP addresses could be found within the Slovak Republic's territory. In the decisions examined, no jurisdiction issues were considered, as the court is not required to consider the possibility of the change of jurisdiction or of applicable law on the ground of an IP address' assignment to internet service providers ('ISPs') from other countries. The foreign IP addresses may also be relevant in the case of different data retention rules (logs storing).

The analysis of the non-anonymised IP addresses also provided interesting information on the question of whether the courts of the Slovak Republic distinguish between static and dynamic, public and private or IPv4 and IPv6 IP addresses, as such classification can serve different purposes in the proceedings.

The categorization of IP addresses as dynamic or static is relevant when determining the time and date of the IP addresses' use (timestamp). To illustrate, if numerous persons had access to the device, to which a specific IP address was assigned, a specific date and time of its use may be necessary to determine the guilty party. However, from the decisions examined, in only a few of them the court mentioned in the judgement's reasoning whether the IP address was dynamic (7 decisions) or static (one decision).

In the decisions themselves, only one decision explicitly stated that the IP address used was public, while no private IP addresses were distinguished. The analysis of the decisions, in which an IP address from the reserved range was contained, has showed that in only a limited number of decisions the fact that such an IP address does not provide internet access was mentioned (*'the IP address does not allow the identification of the computer in the internet network'*). The public IP addresses (determined from the non-anonymised IP addresses) were usually ISP's IP addresses. However, we were also able to find a public IP address that was assigned to a server providing a service (action for damages, where the defendant accessed a specific IP address). Here we can see a possible correlation aspect of public IP addresses. As the court or prosecutor has access to other decisions or case materials, it may be relevant not only to consider the defendant but also other information such as online identifiers present (including IP addresses), e. g. to make use of one of the procedures established in forensic analysis, namely creation of an IP address' profile to find commonalities between different decisions.

As regards the classification of IP addresses to IPv4 and IPv6 IP addresses, none of the decisions examined have provided for such a distinction. This may be the case because all the non-anonymised IP addresses were in the form of an IPv4 address. This distinction, however, will be relevant in connection with privacy and personal data protection issues. These notions are further examined by (von Grafenstein, 2018).

5. Conclusion and future works

Our analysis of the courts' decision-making process provided for interesting results regarding the different dependencies appearing in our dataset. One of the most surprising findings was the conclusion that the more evidence is produced, the more likely it is that the court finds the defendant not guilty or the civil action unsuccessful. In compliance with the in dubio pro reo principle in criminal law, one might assume that the court requires more evidence to determine one's guilt and not innocence.

Additionally, our analysis suggests that the courts' practice of

using IP address as evidence is far from perfect. Even though the IP address primarily identifies the device and not the person, court decisions still assign the IP address to persons. Furthermore, specification of a device is often missing in decisions, as does the deliberation on other persons with possible access to the device beside the defendant. In response to those issues, we believe it would be appropriate to reach some sort of standardisation of the use of IP address as evidence in court proceedings, especially in criminal proceedings.

As regards the second research goal, we can conclude that the IP addresses' anonymisation may still pose a challenge for the courts, as their approach to this issue can differentiate from court to court. There is currently no uniform approach to anonymisation and standardisation we could identify. Possible solutions may include the adoption of recommendations by the Ministry of Justice.

The authors believe that the analysis provides relevant insights as regards forensic analysis and investigation. Its results can be used in the proceedings before the court, e. g. to propose evidence to be executed, which may be helpful to all of the concerned parties. Specifically in criminal proceedings, prosecutors and investigators should focus more on the establishment of the connection between the IP address and the person through the specification of a device to which the IP address was assigned. Direct assignment of the IP address to the person without specific identification of the device may lead to the failure in proving one's case before the court. Moreover, the identification of the four categories of the court's decision-making process may prove helpful for the concerned parties when preparing their strategy for the proceedings, e. g. as regards the evidence to be executed, which may, in turn, lead to making the proceedings more effective.

Acknowledgements

This research is funded by the Slovak Research and development agency (APVV) projects under contract No. APVV-14-0598 and contract No. APVV-17-0561. J.H. would like to gratefully acknowledge the support by European Regional Development Fund (ERDF), 'CyberSecurity, CyberCrime and Critical Information Infrastructures Center of Excellence' (No. CZ.02.1.01/0.0/0.0/16_019/0000822), Czech Republic/European Union. This paper was also created in the framework of the National project IT Academy – Education for the 21st century, which is supported by the European Social Fund and the European Regional Development Fund in the framework of the Operational Programme: Human Resources.

References

Banterle, F., 2018. The interface between data protection and ip law: the case of trade secrets and the database sui generis right in marketing operations, and the ownership of raw data in big data analysis. In: Personal Data in Competition, Consumer Protection and Intellectual Property Law. Springer, pp. 411–443.
Brungs, A., Jamieson, R., 2005. Identification of legal issues for computer forensics. Inf. Syst. Manag. 22, 57–66.
CJEU, 2011. C-70/10 *Scarlet Extended*.
CJEU, 2016. C-582/14 *Breyer*.
Cole, K.A., Gupta, S., Gurugubelli, D., Rogers, M.K., 2015. A review of recent case law related to digital forensics: the current issues. In: 2015 Proceedings of Annual ADFSL COnference on Digital Forensics, pp. 95–104. Security and Law.
Garrie, D., 2014. Digital forensic evidence in the courtroom> understanding content and quality. Northwest. J. Technol. Intellect. Property 12, 121–128.
Goodison, S., Davis, R., Jackson, B., 2015. Digital evidence and the u.s. criminal justice system. URL. https://www.ncjrs.gov/pdffiles1/nij/grants/248770.pdf.
Řřivna, T., Drápal, J., Attacks on the confidentiality, integrity and availability of data and computer systems in the criminal case law of the Czech republic. Digit. Invest. 28, 1–13.
Henseler, H., van Loenhout, S., 2018. Educating judges, prosecutors and lawyers in the use of digital forensic experts. Digit. Invest. 24, S76–S82.
Horsman, G., Findlay, B., James, T., 2019. Developing a 'router examination at scene'standard operating procedure for crime scene investigators in the United Kingdom. Digit. Invest. 28, 152–162.
Ieong, R.S., 2006. Forza–digital forensics investigation framework that incorporate legal issues. Digit. Invest. 3, 29–36.
Issa, H.A., Ismail, M., Aamar, O., 2019. Unauthorized access crime in jordanian law (comparative study). Digit. Invest. 28, 104–111.
James, J., Gladyshev, P., 2016. A survey of mutual legal assistance involving digital evidence. Digit. Invest. 18, 23–32.
Kassambara, A., 2017. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning, vol. 1. STHDA.
Kluwer, W., 2019. Automated legal information system (aspi). URL. https://www.wolterskluwer.sk/sk/system-aspi/o-aspi.c-24.html.
Lee Rodgers, J., Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. Am. Statistician 42, 59–66.
Liles, S., Rogers, M., Hoebich, M., 2009. A survey of the legal issues facing digital forensic experts. In: IFIP International Conference on Digital Forensics. Springer, pp. 267–276.
Lundevall-Unger, P., Tranvik, T., 2010. Ip addresses–just a number? Int. J. Law Info Technol. 19, 53–73.
Manghani, K., 2011. Quality assurance: importance of systems and standard operating procedures. Perspect. Clin. Res. 2, 34.
Menard, S., 2002. Applied Logistic Regression Analysis, vol. 106. Sage.
Montasari, R., 2017. Digital evidence: disclosure and admissibility in the United Kingdom jurisdiction. In: International Conference on Global Security, Safety, and Sustainability. Springer, pp. 42–52.
Nance, K., Ryan, D.J., 2011. Legal aspects of digital forensics: a research agenda. In: Proceedings of the 44th Hawaii International Conference on System Sciences. IEEE, pp. 1–6.
On Legal Co-Operation, E.C., 2016. The use of electronic evidence in civil and administrative law proceedings and its effect on the rule of evidence and modes of proof. european committee on legal co-operation. URL. https://rm.coe.int/1680700298.
Oparnica, G., 2016. Digital evidence and digital forensic education. Digit. Evid. Electron. Signature L. Rev. 13, 143.
Pound, R., 1910. Law in books and law in action. Am. Law Rev. 44, 12–36.
Rogers, M., Scarborough, K., Frakes, K., San Martin, C., 2007. Survey of law enforcement perceptions regarding digital evidence. In: IFIP Intern. Conference on Digital Forensics. Springer, pp. 41–52.
Ryan, D.J., Shpantzer, G., 2002. Legal aspects of digital forensics. URL. http://euro.ecom.cmu.edu/program/law/08-732/Evidence/RyanShpantzer.pdf.
Schwartz, P.M., Solove, D.J., . Problem: Privacy and a New Concept of Personally Identifiable Information. New York University Law Review vol. 86, 1814–1894.
Shukan, A., Abdizhami, A., Ospanova, G., Abdakimova, D., 2019. Crime control in the sphere of information technologies in the republic of Turkey. Digit. Invest. 30, 94–100.
Slay, J., Lin, Y.C., Turnbull, B., Beckett, J., Lin, P., 2009. Towards a formalization of digital forensics. In: IFIP International Conference on Digital Forensics. Springer, pp. 37–47.
Stahl, B., Carroll-Mayer, M., Elizondo, D., Wakunuma, K., Zheng, Y., 2012. Intelligence techniques in computer security and forensics: at the boundaries of ethics and law. In: Computational Intelligence for Privacy and Security. Springer, pp. 237–258.
Stalla-Bourdillon, S., Papadaki, E., Chown, T., 2016. Metadata, traffic data, communications data, service use information... what is the difference? does the difference matter? an interdisciplinary view from the UK. In: Data Protection on the Move. Springer, pp. 437–463.
Tamó-Larrieux, A., 2018. Designing for Privacy and its Legal Framework: Data Protection by Design and Default for the Internet of Things. Springer.
Vincze, E.A., 2016. Challenges in digital forensics. Police Pract. Res. 17, 183–194.
von Grafenstein, M., 2018. The Principle of Purpose Limitation in Data Protection Laws: the Risk-Based Approach, Principles, and Private Standards as Elements for Regulating Innovation. Nomos Verlagsgesellschaft mbH.
Weber, R.H., Heinrich, U.I., 2012. Anonymization. Springer Science & Business Media.
Wegman, J., 2005. Computer forensics: admissibility of evidence in criminal cases. J. Leg. Ethical Regul. Issues (JLERI) 8.
Wong, D.H., 2013. Educating for the future: teaching evidence in the technological age. Digit. Evid. Electron. Signature L. Rev. 10, 16.
WP29, 2007. Opinion 4/2007, on the concept of personal data. URL. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf.