EURASIP Journal on
Information Security

RESEARCH

Open Access

Honeypots and honeynets: issues of privacy

Pavol Sokol[1]*, Jakub Míšek[2] and Martin Husák[3]

Abstract

Honeypots and honeynets are popular tools in the area of network security and network forensics. The deployment and usage of these tools are influenced by a number of technical and legal issues, which need to be carefully considered. In this paper, we outline the privacy issues of honeypots and honeynets with respect to their technical aspects. The paper discusses the legal framework of privacy and legal grounds to data processing. We also discuss the IP address, because by EU law, it is considered personal data. The analysis of legal issues is based on EU law and is supported by discussions on privacy and related issues.

Keywords: Personal data, EU law, Honeypot, Privacy, IP address

1 Introduction

The landscape of cybersecurity threats is continuously evolving and reactive security measures are often not sufficient for protecting information infrastructures. We continuously have to learn about new threats to keep pace with potential attackers.

One of the most popular method of learning about attackers is using *honeypots*. Spitzner defines honeypots as an information system resource whose value lies in an unauthorized or illicit use of that resource [1]. It can also be defined as a computing resource whose value is in being attacked [2]. A honeypot is deliberately allowed to be compromised, and the attack is then analyzed so that we can learn about the methods, procedures, and tools that the attacker used.

It is unquestionable that honeypots increase our understanding of malicious activity in cyberspace. However, we have to keep in mind that there are legal issues regarding honeypots that need to be addressed when deploying one, analysing the captured data, and sharing the results with others. One of the major legal issues is the issue of privacy, which we address in this paper. This issue influences how a honeypot can be deployed, what data they are allowed to collect, and what we can do with the collected data.

To formalize the scope of our work, two research questions are stated:

1. What data are legally allowed to be collected by honeypots?
2. What are the legal conditions for the collection of data and data retention?

In this paper, the authors focus on the European Union (EU) regulations, EU directives, and international agreements. The national legislation of the EU Member States is based on these legal documents (EU directives, international agreements) or alternatively the legal documents are an integral part of the national legislation (EU regulations, international agreements). Therefore, some national legislation may be slightly different from the concept found in the EU law or international law. The aim of this paper is to elaborate on the legal framework of the European Union. We acknowledge that cybersecurity is a global issue where information must be shared across borders and thus there are many legal implications which must be considered within different legal cultures. However, this question is out of the scope of the presented work and will be a subject of future research.

This paper is organized into five sections. The background of honeypots and the related works are discussed in Section 2. This section focuses on previous literature related to the legal aspects of honeypots and honeynets, especially the issue of privacy. Section 3 is the main part

*Correspondence: pavol.sokol@upjs.sk
[1]Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University in Košice, Park Angelinum 9, 040 01 Košice, Slovakia
Full list of author information is available at the end of the article

of this paper and deals with privacy and personal data protection. Section 3.1 focuses on the legal framework of privacy and personal data protection in the EU law. Section 3.2 discusses the basic concepts of personal data protection in the EU. Section 3.3 is focused on the data collected by honeypots and honeynets from the perspective of the EU law. IP addresses as the most important collected data are discussed in Section 3.4. Section 3.5 deals with the legal grounds for data processing and purpose limitation. In Section 4, the paper outlines issues related to privacy, namely network monitoring (Section 4.1) and the publication of results (Section 4.2). Section 5 concludes the paper and outlines the newly opened problems for future research.

2 State of the art

In this section, we present the current state of the art in the discussed topics. First, we introduce honeypots to provide a background in the field. Second, the related work on honeypots and legal issues of honeypots will be presented.

2.1 Background on honeypots

For the purpose of this paper, we classify honeypots according to their level of interaction and purpose. The first classification is based on level of interaction. The *level of interaction* can be defined as the range of possibilities that a honeypot allows an attacker to have. Low-interaction honeypots detect attackers using software emulation of the characteristics of a particular operating system and network services on the host operating system. The advantage of this approach is better control of attacker activities, since the attacker is limited to software running on a host operating system. On the other hand, this approach has a disadvantage: a low-interaction honeypot emulates a service, or a couple of services, but it does not emulate a full operating system. Examples of this type of honeypot are Dionaea [3] and Glastopf [4].

In order to get more information about attackers, their methods, and attacks, we use a complete operating system with all services. This type of honeypot is called a high-interaction honeypot. This type of honeypot aims to give the attacker access to a real operating system, where nothing is emulated or restricted [1]. Examples of this type of honeypot are Sebek [5] and HonSSH [6].

Spitzner suggests the classification of honeypots by *purpose* [1]. There are research honeypots and production honeypots. The research honeypot is designed to gain information about the blackhat community and it does not add any direct value to the organization, which has to protect its information [7]. The main aim here is to get maximum information about the blackhats by giving them full access to penetrate the security system and infiltrate it [8]. A second type of purpose classified honeypot is the production honeypot, used within an organization's environment to protect the organization and help mitigate risk [7]. An example of the production honeypot is a honeypot which captures, collects, and analyzes malware for anti-virus systems, intrusion detection system signatures, etc.

Honeynet extends the concept of a single honeypot to a highly controlled network of honeypots [9]. A honeynet is composed of four core elements [7, 10]:

- *Data control*—monitors and logs all of the activities of an attacker within the honeynet
- *Data capture*—controls and contains the activity of an attacker
- *Data collection*—stores all captured data in one central location
- *Data analysis*—the ability of the honeynet to analyze the data being collected from it

The deployment and usage of honeypots bring many benefits, e.g., the possibility of discovering new forms of attacks. In addition, low-interaction honeypots are easy to deploy, undemanding resource-wise, and simple to use [7]. On the other hand, a number of issues need to be addressed during the deployment and usage. The most frequent problems are [11]:

- *Inaccurate results*—in some cases, the data obtained from the honeypots lead to poor results, due to a limited amount of data
- *Discovery and fingerprinting*—the attackers can detect the honeypots
- *Risk of takeover*—the honeypot may be used to attack against the real (non-honeypot) systems

The quantity and quality of the data collected from honeypots are one of the problems associated with their usage. This problem is closely linked to the *issue of privacy*. It represents one of the most significant concepts in the field of law, and it was set forth in Article 8 of the European Convention on Human Rights. Privacy can be defined as the right to be left alone and to have a private life is [7]. It can also be defined as the right of a person to be free from unwarranted publicity [9].

This includes some individual privacy, such as the privacy of the home and office, the protection of physical integrity, and also the privacy of communications (telephone calls, chats, emails etc.). Therefore, the primary motivation for writing this paper is the fact that an administrator has to take into account the issue of privacy and related issues in the process of data collection. The failure of an administrator to meet that responsibility leaves them open to a lawsuit for any disruption of privacy and resulting damages.

2.2 Related works

The papers dealing with the legal aspects of honeypots and honeynets focus on three fundamental legal issues of the deployment and usage of honeypots: privacy, liability, and entrapment. We discuss them in more detail below. They only deal with privacy in the context of honeypots only. Most of papers focus on legal issues from the *US law perspective*.

Mokube and Adams [11] focus on the aspects of the deployment and usage of honeypots in the USA in general. One of these aspects is the legal issues. According to him, the laws might restrict the right to monitor users on a system. Scottberg [12] outlines the privacy issues of the attackers' files, which are uploaded to the servers by attackers. According to him, these files are not protected. Salgado [2] outlines the legal framework of the usage of honeypots. He recommends taking into account the laws that restrict the monitoring of users' activities. Salgado extends his analysis in the paper [13]. An important analysis of the legal aspects of honeypots from the US perspective is presented by Spitzner [2]. He discusses the same legal issues as the previous papers. In issues of privacy, he distinguishes two types of information being collected: transactional and content.

There are papers that at least outline the legal issues of honeypots from the *perspective of the EU law*. For example, Dornseif et al. [14] focus on legal issues of the usage of honeypots in the context of German laws. Sokol [15] focuses on the legal issues of honeynet generations. He discusses in particular the privacy and liability issues in each generation.

The abovementioned papers deal directly with honeypots. There are a number of papers focusing on the legal aspects in related fields, such as digital forensics and cybersecurity.

Since honeypots belong to network forensics tools, the legal aspects of *digital forensics* are relevant. Nance et al. [16] introduce a preliminary research hierarchy for legal issues associated with digital forensics. The topic discussed in their paper includes property law, constitutional law, tort law, contract law, cybercrime, criminal procedure, evidence law, and cyberwar. Another interesting paper is about legal and technical issues of Internet forensics [17]. This paper provides a combined approach on the major issues pertaining to the investigation of cybercrimes and the deployment of Internet forensics techniques. It discusses major issues from a technical and legal perspective, and it provides general directions on how these issues can be tackled. The paper also discusses the implications of data mining techniques and the issue of privacy protection with regard to the use of forensics methods.

Another related field of research is *cybersecurity*. Burstein [18] focuses on issues related to cybersecurity research, especially running infected hosts, testbeds, non-isolated hosts, publishing results, etc. Another very interesting paper in this field relates to the legal issues surrounding monitoring during network research [19]. There, Sicker et al. focus on several US laws that prohibit or restrict network monitoring and the sharing of records of network activity.

3 Privacy and personal data protection

In this section, we discuss selected aspects of privacy and data protection in the area of honeypots. First, we outline framework of privacy in the EU law. Then, we discuss privacy issues concerning data collected by honeypots, IP addresses, and data processing.

3.1 Legal framework of privacy and personal data protection in EU law

This section provides an overview of the most important privacy regulations in the EU that are applicable to honeypots. The EU legal framework, applicable to honeypots and honeynets, consists of the following *legal instruments*:

The primary regulatory instrument, or the lex generalis, of the personal data protection system is currently EU Directive 95/46/EC focused on the protection of individuals with regard to the processing of personal data and on the free movement of such data. It is commonly known as the *EU Data Protection Directive*. It ensures an equivalent level of protection of fundamental rights and freedoms, and in particular, the right to privacy, with respect to the processing of personal data. It also ensures the free movement of such data within the EU and sets rules for transborder data flow outside of the EU. The directive will be replaced by EU Regulation No. 2016/679, the General Data Protection Regulation (GDPR), which comes into force on 25 May 2018. GDPR is based on the same principles as the Directive, and thus, it will not change the basic premises of the data protection system. However, it adds a number of duties for the data controller (the person who processes the data and sets the purpose of the processing) and rights for the data subjects (an identifiable person, whose data is processed).

EU Directive 2002/58/EC focuses on the processing of personal data and the protection of privacy in the sector of electronic communications. This directive is commonly known as the EU Directive on privacy and electronic communications (*e-Privacy Directive*). This directive is lex specialis, and it specifically regulates privacy and personal data protection when it comes to electronic communications. It has to be used and interpreted in accordance with the general act, which is the Data Protection Directive, or soon the GDPR. This directive covers and harmonizes certain issues of privacy in electronic communications. Some of them are universally binding, e.g., preserving the confidentiality of communication and specific regulation of cookies; however, others only regulate operations of electronic communications providers, e.g., the storing of traffic and location data. On January 2017, the European Commision introduced a proposal for a new e-privacy Regulation. It will replace the current Directive, and thus, the material scope of it will probably stay similar. One of interesting novelties is an explicit notion of machine-to-machine communication falling within the scope of the Regulation. It is too early now to draw any specific conclusions, because the proposal must go through the whole legislative process.

The last relevant piece of legislation is the EU Directive on the security of network and information systems (2016/1148/EU; the "NIS Directive") which was enacted on 6 July 2016. The main purpose of the NIS Directive is to harmonize cyber security infrastructures of the member states so they can easily share information concerning cyber security incidents. Therefore, the NIS Directive may serve as a basis for national legislation, which will put information sharing duties upon certain honeypot operators. However, the directive explicitly states in Art. 2 that any processing of personal data pursuant to it must be carried out in accordance with legal acts on data protection.

3.2 Basic concepts of personal data protection

In this section, we present several basic concepts of the *European personal data protection system*, which are relevant for honeypots and their functions and data processing. The data protection system is based on a principle of preventing privacy harm [20]. To achieve this, the Data Protection Directive incorporates a very broad definition of "personal data," so the highest possible number of persons can be considered "data controllers." The most important duty of the controller is to process personal data only for legitimate and legal purposes and based on a legitimate legal ground. All this combined can ensure a high level of protection, as required by the recital 10 of the Personal Data Protection Directive and the Court of Justice of European Union (CJEU) in recent cases concerning personal data protection, e.g. Google Spain case C-131/12, Rynes case C-212/13, and Schrems Case C-362/14.

Personal data is defined in the Art. 2 letter a) of the Personal Data Protection directive as follows: "any information relating to an identified or identifiable natural person; an identifiable person is one who can be identified, directly or indirectly." The most relevant part of the definition is the notion of indirect identifiability. It means that any information, which can be used in the right context for the identification of a person ("data subject") is personal data, even though the information in itself (outside the right context) does not directly identify the data subject [21]. This approach, which is supported by the CJEU, since it is necessary for assuring a high level of protection, leads to a situation where almost any information could be personal data. Thus, operators of honeypots and honeynets should be aware of this situation.

A purpose is the cornerstone of every personal data processing. It is set by the data controller and all that happens to the data during its life cycle is connected with the set purpose. Personal data can be processed only in order to achieve the declared purpose, which has to be conveyed to the data subject. The personal data can also be retained only for a time period that is necessary for fulfilling that purpose. This principle is called "the purpose limitation," it is grounded in Art. 6 para. 1 letter b), and it applies also on the legal grounds for processing. Once the purpose changes, or the current legal ground can no longer be used or relied on, the data controller has to find another legal ground or cease the data processing.

The Data Protection Directive recognizes in the Art. 7 several legal grounds for data processing, from which the following are relevant for the case of honeypots and honeynets:

- The data subject has unambiguously given their consent (letter a))
- The processing is necessary to comply with a legal obligation to which the controller is subject (letter c))
- The processing is necessary for the purposes of legitimate interests pursued by the controller or by a third party or parties to whom the data are disclosed, except cases where such interests are overridden by the interests for fundamental rights and freedoms of the data subject (letter f))

The Data Protection Directive sets four conditions for the validity of consent which must be met. It has to be freely given, specific, informed, and unambiguous [22]. However, there are several practical problems with this concept both on the side of the data subject (e.g., no one reads the terms and conditions and almost nobody can understand them; therefore, it is quite questionable whether the given consent is in fact informed) [23] and on the side of data controller (e.g., it is technically almost impossible to obtain legally valid consent of data subjects whose personal data are processed in the course of honeypot and honeynet operation). Apart from the legal grounds for processing because of the necessity arising from a legal duty, the Data Protection Directive offers legal grounds for processing of personal data for the legitimate interests of the data controller or a third party. This legal duty must be grounded in a public law norm. For the legal duties of the data controller, which arising from private law, the provision of Art. 7 letter b) of the data protection directive is applicable. Working Party 29 elaborated on this issue in its opinion No. 6/2014 [24], which can be summed up by stating that personal data can be processed for the legitimate interest of the data controller or the third party, as long as it is proportional with the impact on the right of privacy of the data subject.

3.3 Collected data

As described in the previous section, almost any data collected by honeypots might be considered personal data. The first aspect of privacy issues within honeypots and honeynets is the type of data that is being collected. There are two general categories:

- The contents of communications
- Information to establish communication

The first type of collected data, the *contents of communications* (*content data*), is regulated by the EU Directive on privacy and electronic communications. According to Article 2 a) of the e-Privacy Directive, communication (content data) means "any information exchanged or conveyed between a finite number of parties by means of a publicly available electronic communications service." Examples of content data are the bodies of email messages, file contents, full packets captured on a network segment, reconstructed content of interactive sessions (e.g., commands executed in a shell account, typed passwords), etc. Apart from harmonized European law, the demand for communication confidentiality is included in national legal regulation and is protected by civil as well as criminal law.

The extent of the collected content data records is related to the honeypot's level of interaction. Low-interaction honeypots capture and collect smaller amounts of content data records than medium-interaction and high-interaction honeypots.

The second type of collected information records is the *information to establish communication* (no-content data, transactional data, also known as *metadata*). These are mostly traffic and location data, which are defined in the EU Directive on Privacy and Electronic Communications, as follows:

1. *Traffic data*—any data processed for the purpose of conveying a communication on an electronic communications network or for the billing thereof (Article 6 of that Directive)
2. *Location data*—data processed in an electronic communications network, indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service (Article 7 of that Directive);

Examples of transactional data are IP addresses, network ports, network protocols, account names, email header information, time, date, website URLs, etc.

The categories of transactional data retained in honeypots include:

1. Data necessary to trace and identify the *source* and *destination of a communication*, for example the IP address and domain name
2. Data necessary to identify the *date, time, and duration of a communication* (e.g., timestamp)
3. Data necessary to identify the *type of communication*, for example an Internet protocol (e.g., ftp, ssh, samba)
4. Data necessary to identify *the users' communication equipment* or what purports to be their equipment, for example the operating system

From the perspective of honeypots, the IP address, timestamp, and Internet protocol are data collected in all honeypots. Due to the abovementioned broad definition of personal data, all of this should be considered personal data within the scope of the Data Protection Directive.

3.4 IP addresses

In this section, we argue that *IP addresses are personal data* in the meaning of the Data Protection Directive. As stated before, the IP address is a piece of information necessary to trace and identify the source of a communication. According to our opinion, it is the most important piece of information in any subsequent analysis.

An IP address is connected with a specific device [25]. However, in many cases, we can assume a strong connection between the device and its user. That is the case of smart phones, tablets, and other smart handheld devices, as well as personal computers. IP addresses are used by electronic communications service providers to help identify a subscriber [26]. IP addresses are also collected and stored by electronic communications providers for the purpose of a possible criminal investigation. This is done in the course of data retention duty, which is still present in several member states although the Data retention directive 2006/24/EC was nulled by the CJEU. We can see in this example that IP addresses are used as information which leads to the identification of a person. Therefore, it counts as indirectly identifying personal data.

This view is supported both by the Data Protection Directive Article 29 Working Party, which considers IP addresses to constitute personal data within the meaning of Article 2 a) of the EU Data Protection Directive [21] and the CJEU. The CJEU dealt with IP addresses in the case Scarlet Extended SA vs. Socit belge des auteurs compositeurs et diteurs (SABAM) (C-70-10). In this case, the CJEU stated in Section 51 that the monitoring of the behavior of Internet users and any further collection of their IP addresses amounts to an interference with their rights to respect for their private life and their correspondence, since IP addresses are personal data.

In this respect, the prejudicial question of the Federal Government of the Federal Court ("BGH") about IP addresses is quite crucial. In what is now known as the Breyer case (C-582/14), the BGH filed a preliminary reference to the CJEU on whether dynamic IP addresses are at all considered personal data, protected by the European data protection law, even if no further information on the identity of the terminal holder is available. In his opinion, the Advocate General stated that "an IP address stored by a service provider in connection with access to its web page constitutes personal data for that service provider, insofar as an Internet service provider has available additional data which make it possible to identify the data subject" [27]. This case is very similar to data collection in honeypots. That would mean that IP addresses are not personal data in the situation of honeypot and honeynet operators, because the particular natural person is not identifiable by the means the operator has at their disposal. Furthermore, in most situations, the attack is carried out by a machine, not a human. In this case, an identification of the natural person is fairly difficult. However, the final ruling stated in paragraph 49 "dynamic IP address...when a person accesses a website that the provider makes accessible to the public constitutes personal data within the meaning of that provision, in relation to that provider, where the latter has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person" [28]. This provision softened the objective approach to personal data definition a little. However the "legal means" which the court mentioned might be for example just a possibility to hand the data over to the police which has then access to data retention data.

Therefore, in our opinion, it is safer to consider IP addresses personal regardless of what other information the operator has. There are three reasons for that. First, it is the basic preventive principle of the personal data protection system, which regulates the amount of collected data, so it cannot be connected and misused. Second, the opinion of the Advocate General is not binding for the CJEU and should the court keep the line of its previous decisions, it might decide on the matter more strictly. Third, even though in a number of cases the IP addresses can be connected only to a device and not a human being (e.g. the Internet of things), there is not an easy way for the honeypot operator to distinguish them [29].

3.5 Legal grounds to process data and purpose limitation

IP addresses collected during the operation of honeypots and honeynets can be personal data of either the operator's customers or third persons, whose devices are used for the attack. The customers can provide consent for the personal data processing, but that is not the case for the third persons. Furthermore, it is advisable to rely on a different legal ground for processing than for consent, when it is available and applicable [23]. The legal ground must be chosen according to the purpose of the processing.

The following may be considered a *relevant purpose* of personal data processing within honeypots and honeynets:

- For *production honeypots*—safeguarding the security of the service
- For *research honeypots*—research and prevention of future threats

In the first case, the data controller can rely on their legitimate interest in the cybersecurity of his network. The possible harm of privacy for the data subject (those whose IP addresses are processed) is very little. Therefore, they can process personal data in accordance with Art. 7 letter f) of the Data Protection Directive. Furthermore, this processing is also in accordance with the legitimate interest of the owners whose devices were used for the attack, since this processing might help to solve their unfortunate situation.

In the second case, the situation is more complicated. The legitimate interest of the controller might be a promotion of cyber security and a right to carry out their business properly. These interests must be proportionate with the right of data subjects for privacy protection in the light of the possible harm done by the processing. Since the possible harm is quite low, we are convinced that the legal ground for processing established in Art. 7 letter f) should be applicable in the case of research honeypots as well.

Furthermore, the data controller must consider an adequate period of time for the retention of the collected data. It is, yet again, connected with the purpose of the processing. As was mentioned earlier, the data controller can only hold the data only a necessary period.

In the case of production honeypots, the data should be erased periodically after a shorter period of time (e.g., one month) or once the security incident is resolved. In the case of research honeypots, it could be a longer time, but it must not exceed proportionality of the Art. 7 letter f) legal ground. Should the data controller wish to keep the data longer, they would have to obtain consent from the data subjects.

Finally, regardless of whether it is a case of a production or research honeypot, the honeypot operator might have a legal duty to share information about cybersecurity incidents based on, for example, the NIS Directive. In such cases, the data can be processed (and transferred) in accordance with the provision of Art. 7 letter c), as mentioned above.

In situations when the data transfer is not prescribed by the law, it is necessary to rely on different provisions. In case of data transfer within the borders of the European Union, the European Free Trade Area, and countries with an adequate level of protection, that would be again the Art. 7 letter f) legal ground. Article 25 of the Data Protection Directive enables the Commission to promulgate an adequacy decision, which states that the country in question has an adequate level of protection of personal data adequate to that of the EU. We argue that in this case, the legitimate interest on the processing might be the interest of users of communication networks because the sharing of information improves the security of the whole network ecosystem. It is true that this interest may seem quite vague, but as long as it is proportionate with the rights of the data subject, it is legal. The proportionality is, in our point of view, ensured by the fact that IP addresses in themselves do not impose too much of a privacy threat. In case of information sharing to partners seated in other countries, general rules on transborder personal data transfer apply [30].

4 Other privacy issues

4.1 Network monitoring

Another set of issues associated with the daily functioning of honeypots and the realities of their operation is related to the very nature of honeypots in the area of research. A *monitoring network* may contribute to its improved security or valuable research output, whether we are talking in terms of production or research honeypots. There are several questions to deal with, namely the proportionality of the invasion of data for research purposes or for security. Monitoring every single packet, as has been shown elsewhere [19], may lead to considering this kind of situation a threat in itself (the question on who will control the guards), although we point out that courts will look at industry practices.

Apart from that, the EU prefers a universal legal framework, as opposed to specific industry practices or sectorial regulation, as is the case in the USA. It would be hard to define precisely which jurisdictions would bring what kind of decisions and how, due to a lack of precedent in most jurisdictions. However, we should remember that the legal framework on fundamental rights is strict and that this legislation is transparent according to its historical development.

Any research would have to comply with the existing legislation and it cannot be seen as legally entitled for such review or control, because these powers typically belong to public authorities and cannot be delegated to private entities, as this would entirely undermine the philosophy of data protection. Monitoring networks must meet some limits and adhere to valid standards which will not violate valid legislation or pose a threat to society from the risk of abuse of these facilities. Legislators should avoid an unbalanced exercise of security measures, which poses a threat to civil society and privacy.

Although we were discussing issues related to clashes of values and their legal quality in previous sections, we have not put focus on policy issues related to implied risks, which are related to the volume of data. If we take a look at this in the light of data retention, the retention of data for considerably long periods of time leads to risk exposure, which will pose a long-term threat to privacy and security to guarantee privacy rights in practice.

If data are stored, if all traffic is stored or monitored, then a leak of these data is a probable threat and a theft of these data is more probable the longer the period of time, because nothing is perfectly secure on the Internet and risks will become more salient as the time for their emergence and chance to occur increases. By trying to be in a state of higher security, we are actually risking more by creating implied insecurity. Thus, efforts to maintain more secure societies may lead to societies which will have to face environments with higher risks and fewer factual securities, because their data will be exposed to these risks for longer periods of time. This means that protection cannot cross a certain line; it must be proportionate. These conclusions could be summarized as follows: if there is an absolute, large amount of information stored, the risk of it being stolen grows with time and volume. In other words, the Internet is not perfect and security systems can be invaded, attacked, and penetrated successfully. This argument is quantitative in its essence.

A qualitative argument could be derived from different situations. What if a honeypot worked as a support facility to a chat server or another electronic service and research and security would require specific data? This leads to another problem which arises in cases of secrecy of correspondence. Even a network security and public order provisions should not deprive us of this right.

If a honeypot is deployed within a chat service or a similar service, the secrecy of messages has to be guaranteed. In case that a researcher identifies, using a research honeypot, the possibility that there is some kind of a suspicious activity or pattern, this implies that they will use these data for analysis. This also includes text messages. However, if we deal with content data, a second-order error may occur: the privacy of someone who was not conducting acts of a malicious character will be disclosed to a third person, thus violating data protection and privacy rights. From a procedural perspective, this produces a legal problem on the grounds of criminal and constitutional law, because interventions into privacy, such as wiretapping or other forms of monitoring, would proceed without a court order. An invasion of privacy by an administrator or researcher depends on the public law provisions which authorize only those actions which do not intrude on privacy without legitimate reasons recognized and defined by the law. Under any other conditions, an invasion of privacy, which is such a serious intervention that a court order is required, cannot take place and the administrators have to comply with and respect privacy. This defines the limits of research or security functions administered by honeypot administrators.

4.2 Publication of the results

Publication of results is related to the privacy issues outlined in the subsection relating to data capture. One of the important problems within this issue is the sharing and publishing of network traces. The scientific motivations for sharing these data are compelling: common datasets can provide meaningful comparisons between competing research approaches; simulated data are inadequate for some uses; and existing datasets may not reflect present-day threats or traffic characteristics [31]. In this aspect, it is necessary to mention the anonymization issue. Before presenting research data, it is necessary to anonymize these data. Network trace anonymization is an active area of research in the security community, as shown by the ongoing development of anonymization methods and the releases of network data that they enable [32]. Since the results contain personal data, their publication would constitute a new personal data processing with a new purpose and legal ground. Since this might be quite problematic, it is recommended to publish only anonymized version of the results.

The publication of results also has the potential to harm an organization's reputation by revealing network details that the institution would prefer to keep secret. A strictly legal concern that this raises is the potential for a breach of contract [33]. The possibility that a publication will reveal details about a honeypot or a production network also raises liability issues. Honeynet administrators should also consider whether the papers or datasets that they publish could reveal information that could help adversaries attack the honeynet or production network of an organization. Publishing datasets is likely to pose a greater risk to a production network than a paper; therefore, data releases may deserve a more careful vetting by IT officers than papers do [33]. Another aspect of liability is the fact that the publication of results merely provides information that might help another person commit cybercrime [34].

5 Conclusions

The legal aspects of honeypots and honeynets are a fascinating research topic. In this paper, we have discussed the aspects of privacy and personal data protection. The paper has outlined the concept of personal data protection in the EU law and focused on the issues of the purpose of data processing, the legal grounds for data processing, and the retention of data. The paper has also discussed issues related to privacy, such as network monitoring and the publication of results.

First, we investigated the data collected by honeypots from the perspective of the EU law. In data collection, it is necessary to distinguish content data (contents of communications) and transactional data (information for establishing communication). It is also crucial to identify a relevant lawful purpose of personal data processing and choose a correct legal ground for such a processing.

Second, we studied the legal conditions for the collection of data and data retention. Administrators of honeypots and honeynets, who are seen as personal data controllers in the eyes of the law, because IP addresses are considered personal data within EU law, can rely on the legal ground of legitimate interest to collect and process personal. In production honeypots, the legitimate interest lies in safeguarding the security of the service. In our opinion, the legal ground of legitimate interest is applicable also for research honeypots. An adequate length of retention of personal data is also an important issue for the processing of honeypot data, since the data minimization principle applies. In situations where a researcher wants to publish data collected by honeypots and honeynets, anonymization is needed.

The conclusions of this paper open issues that need to be addressed in the context of future research. In connection with the fact that IP addresses are personal data, it is necessary to discuss them in more detail and propose an anonymization technique for the collected data. Other newly opened research questions are closely linked to international cooperation and the cooperation with private and public authorities. In these cases, it is needed to closely discuss the issue of cross-border transmission of data.

Acknowledgements
We would like to express special thanks to our colleagues from the Czech chapter of The Honeynet Project for their comments and valued input. The research was supported by the Slovak APVV project under the contract No. APVV-14-0598 and the Masaryk University grant no. MUNI/A/1191/2016.

Authors' contributions
PS outlined the concept of legal research in the field of honeypots and the concept of this paper. He has been involved in drafting the manuscript. JM focused on the basic concepts of personal data and collected the data and the legal analysis of IP addresses. MH participated in technical issues of honeypots and has been involved in the critical revisions of the draft. All authors have read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Institute of Computer Science, Faculty of Science, Pavol Jozef Šafárik University in Košice, Park Angelinum 9, 040 01 Košice, Slovakia. [2]Institute of

Law and Technology, Faculty of Law, Masaryk University, Veverí 158/70, Brno, Czech Republic. [3]Institute of Computer Science, Masaryk University, Botanická 68A, Brno, Czech Republic.

Received: 2 October 2016 Accepted: 16 February 2017
Published online: 28 February 2017

References

1. L Spitzner, *Honeypots: tracking hackers*. (Addison-Wesley Reading, Boston, 2003)
2. L Spitzner, The honeynet project: trapping the hackers. IEEE Security and Privacy. **1**(2), 15–23 (2003)
3. Dionaea Project. https://github.com/rep/dionaea. Accessed 20 Aug 2016
4. Glastopf Project. http://mushmush.org/. Accessed 20 Aug 2016
5. Sebek Project. https://projects.honeynet.org/sebek/. Accessed 20 Aug 2016
6. HonSSH Project. https://github.com/tnich/honssh/wiki. Accessed 20 Aug 2016
7. A Mairh, D Barik, K Verma, D Jena, in *Proceedings of the 2011 International Conference on Communication, Computing & Security. ICCCS '11*. Honeypot in network security: a survey (ACM, New York, 2011), pp. 600–605
8. N Provos, *et al*, in *USENIX Security Symposium*. A virtual honeypot framework, vol. 173 (The USENIX Association, Berkeley, 2004), pp. 1–14
9. FH Abbasi, R Harris, in *Telecommunication Networks and Applications Conference (ATNAC), 2009 Australasian*. Experiences with a generation III virtual honeynet (IEEE, Piscataway, 2009), pp. 1–6
10. L Spitzner, in *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*. Honeypots: Catching the insider threat (IEEE, Piscataway, 2003), pp. 170–179
11. I Mokube, M Adams, in *Proceedings of the 45th Annual Southeast Regional Conference. ACM-SE 45*. Honeypots: concepts, approaches, and challenges (ACM, New York, 2007), pp. 321–326
12. B Scottberg, W Yurcik, D Doss, in *Technology and Society, 2002.(ISTAS'02). 2002 International Symposium On*. Internet honeypots: Protection or entrapment?, (IEEE, Piscataway, 2002), pp. 387–391
13. The Honeynet Project, *Know your enemy: learning about security Threats*. (Addison-Wesley Reading, Boston, 2004), p. 800
14. M Dornseif, FC Gärtner, T Holz, Vulnerability assessment using honeypots. Praxis der Informationsverarbeitung und Kommunikation. **27**(4), 195–201 (2004)
15. P Sokol, in *Electronics, Computers and Artificial Intelligence (ECAI), 2014 6th International Conference On*. Legal issues of honeynet's generations (IEEE, Piscataway, 2014), pp. 63–69
16. K Nance, DJ Ryan, in *System Sciences (HICSS), 2011 44th Hawaii International Conference On*. Legal aspects of digital forensics: a research agenda (IEEE, Piscataway, 2011), pp. 1–6
17. M Karyda, L Mitrou, in *IEEE Second International Workshop on Digital Forensics and Incident Analysis (WDFIA 2007)*. Internet forensics: legal and technical issues (IEEE, Piscataway, 2007), pp. 3–12
18. AJ Burstein, Conducting cybersecurity research legally and ethically. LEET. **8**, 1–8 (2008)
19. DC Sicker, P Ohm, D Grunwald, in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*. Legal issues surrounding monitoring during network research (ACM, New York, 2007), pp. 141–148
20. R Gellert, Understanding data protection as risk regulation. J. Int. Law. **18**(11), 3–16 (2015)
21. Article 29 data protection working party: opinion no. 4/2007 on the concept of personal data, from 20th June 2007. 01248/07/EN. WP 136. Brussels, Belgium (2007). Article 29 Data Protection Working Party
22. Article 29 data protection working party: opinion no. 15/2011 on the definition of consent, from 13th July 2011. 01197/11/EN. WP187. Brussels, Belgium (2011). Article 29 Data Protection Working Party
23. J Míšek, Consent to personal data processing—the Panacea or the dead end? Masaryk Univ J Law Tech. **8**(1), 69–83 (2014)
24. Article 29 data protection working party: opinion no. 6/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, from 9th April 2014. 844/14/EN. WP217. Brussels, Belgium (2014). Article 29 Data Protection Working Party
25. JJ McIntyre, Balancing expectations of online privacy: why internet protocol (IP) addresses should be protected as personally identifiable information. DePaul Law Review. **60**(3), 895–948 (2011)
26. AV Litvinov, Data protective directive as applied to internet protocol (IP) addresses: uniting the perspective of the European Commission with the jurisprudence of Member States. Geo. Wash. Int'l L. Rev. **45**, 579–610 (2013)
27. Advocate general: paragraph 78 of the opinion of advocate general Campos Sánchez-Bordona Delivered on 12 May 2016 on the case C582/14 Patrick Breyer V Bundesrepublik Deutschland. Luxembourg (2016). Advocate General
28. Case C-582/14 - Judgment of the court (second chamber) of 19 October 2016 (request for a Preliminary Ruling from the Bundesgerichtshof – Germany) – Patrick Breyer V Bundesrepublik Deutschland (2016)
29. Míš,ek, Haraš,ta, IP adresy v kybernetické bezpečnosti. Revue pro právo a technologie. **6**(12), 21–42 (2015)
30. M Zalnieriute, Transborder data flows and data privacy law. Comput Law Secur Rev Int J Tech Law Pract. **1**(30), 104–108 (2014)
31. J Che, Q He, K Ye, D Huang, in *High Performance Computing and Applications: Second International Conference, HPCA 2009, Shanghai, China, August 10-12, 2009, Revised Selected Papers*, ed. by W Zhang, Z Chen, CC Douglas, and W Tong. Performance combinative evaluation of typical virtual machine monitors (Springer, Berlin, Heidelberg, 2010), pp. 96–101
32. P Defibaugh-Chavez, R Veeraghattam, M Kannappa, S Mukkamala, A Sung, in *2006 IEEE Information Assurance Workshop*. Network based detection of virtual environments and low interaction honeypots (IEEE, 2006), pp. 283–289
33. R Sira, Network forensics analysis tools: an overview of an emerging technology. GSEC, version **1**, 1–10 (2003)
34. CL Brown, *Computer evidence: collection and preservation*. (Nelson Education, Toronto, 2010)