Anonymization of Measurement and Monitoring Data: Requirements and Solutions

Fabian Haibl[1] and Falko Dressler[2]

[1] Computer Networks and Internet, Wilhelm-Schickard-Institute for Computer Science, University of Tübingen

[2] Autonomic Networking Group, Dept. of Computer Science 7, University of Erlangen

It is commonly agreed that measurement and monitoring of communication networks are necessary steps to obtain knowledge about the current state and behavior of the network. Such information can be used for accounting or traffic engineering solutions as well as for network security approaches such as attack detection and prevention. At the same time, as more efficient monitoring tools are developed, conflicts with the protection of data privacy arise. In this article, we provide a study of the current legal situation in Germany. As can be seen, there are many flaws in the German law concerning the appropriateness of particular monitoring solutions. We also discuss possible approaches for anonymization of measurement and monitoring data and evaluate their applicability based on the legal foundations.

1. Introduction and Motivation

1.1 Monitoring and Measurement in the Internet

In the age of our information society, computer networks increasingly adopt the role of a base infrastructure. This growing importance is accompanied by new requirements: the economic value is to be skimmed, performance and permanent availability need to be assured, and the security of the network must be guaranteed. These goals can be approached by applications such as accounting, traffic engineering and intrusion detection (Anagnostakis, Ioannidis et al. 2002). The basis of all these applications is a network monitoring infrastructure. Obviously, this requires the gathering, processing, and storage of personal data, so conflicts with the protection of data privacy arise.

The objective of this paper is to state requirements for the collection of network traces and IDS signatures, to present the current legal position in Germany, to develop possibilities of anonymization, and to introduce a rule based anonymization module.

At first the importance of data privacy is illustrated in a historical review. Thereafter, the legal situation in Germany is described. The capability to relate information to a natural person as the prerequisite for the applicability of any privacy related regulation is stated and determined how far an IP address is suitable to create such a link to a person. The applicable German law is determined with regard to monitoring and intrusion detection. After deciding the processing of personal data for various reasons admissible, the modalities are deduced from the principles of avoidance and thriftiness in processing personal data. Thus, the collected data has to be pseudonymized, anonymized, or deleted as soon as possible if the intended purpose can still be accomplished (Pang and Paxson 2003).

Subsequent to the legal expertise, the design and discussion of various anonymization algorithms take place. These comprise the retainment of the IP addresses in a shortened form or of their hash value, their encryption, or their mapping by an injective and prefix preserving function.

Finally a framework is introduced which parameterizes and applies anonymization algorithms depending on addresses and ports.

1.2 Comparability of Traces and Results

In the networking community, many research papers and online available documents discuss and evaluate monitoring data. Often, the quality of these documents suffers from impossible comparability and insufficient comprehensibility.

Two cases have to be distinguished: The analysis described in a research paper is based on online available monitoring data or packet traces. In this case, the data were usually anonymized and somehow falsified. Without knowledge about the used anonymization rules, the experiments cannot be repeated and, therefore, the results cannot be verified. The other possibility is that used packet traces are not made publicly available to protect the privacy of observed Internet connections. In this case, nothing can be done to evaluate the measurement results (Paxson, Mahdavi et al. 1998).

This problem was identified by several research groups. For example, CAIDA (Cooperative Association for Internet Data Analysis) is working on an archive of measurement data and packets traces. The CAIDA website indexes the data collections; while many are publicly available, others are not. CAIDA personnel work continuously to expand the availability and accessibility of the data. Another example is the following call from PAM 2006 (Passive and Active Measurement Conference). PAM steering committee believes that releasing measurement data allows for

better science to be conducted in the field of network measurement. Therefore, for the first time, an award will be given at PAM 2006 for the best paper based on a new dataset that the authors are releasing for community use in subsequent research. To qualify, a paper must significantly utilize a dataset that has been collected for the work presented in the paper. Further, the dataset must be freely available to any researcher. Standardized anonymization are also be investigated by the IETF (Internet Engineering Task Force).

In summary, it can be said that two key requirements must be fulfilled for each provided data set in the area of measurement and monitoring: comparability, i.e. repeatability of provided analyzes, and comprehensibility, i.e. provision of complete information about the monitoring and anonymization process.

1.3 Application Scenarios

Measurement and monitoring data are needed for many purposes. Traffic engineering and analysis is only part of this research domain. Obviously, the described key requirements are directly applicable in order to allow the test and verification of different methods and protocols developed for Internet routing. Another application domain is network security. In this research area, packet traces and signatures are needed for portable attack detection mechanisms. Signatures are commonly used in intrusion detection systems to identify worms and other attacks. Such signatures must be, on the one hand side, available and, secondly, tested on previously collected packet traces. The need for anonymization and privacy is discussed in the following section.

2. Privacy and Legislation

2.1 The Need for Privacy – An Historical Overview

"The objective of data privacy is to protect man from risks resulting from disadvantageous aftermath of data processing" (Schaar, 2004).

At the end of the nineteenth century the basic idea of data privacy arose. In the pre-industrial American society, it was considered unseemly to intrude on private affairs. The "right to be let alone" (Brandeis, 1890) as a space of individual freedom and self-fulfillment was in need for protection particularly with the upcoming of modern mass media. The further development in the United States and Europe diverged. While data protection in Europe addressed nearly any processing of personal data, the US notion of privacy dealt with the protection against infringements of private affairs and collection of personal data by the state by granting informational claims to the citizen regarding the stored data (Kahlert, 2004). With the increasing use of computers as an aid for the automatic

processing of data the conflict about privacy culminated when a committee proposed a central database to record all available data about each US citizen. With the vision of George Orwell's Big Brother in 1984 in mind, the database was not established for the good of privacy (Tinnefeld, 2005).

In Germany, the consciousness about data protection arose after the Second World War with the automation of data processing (Kahlert, 2004). Besides only partial legal regulations, the Federal Constitutional Court (≈ Supreme Court) deduced a constitutional claim of the individual for an unimpeachable private sphere.

Information has always to be regarded with its situational context. The view of its accompanying circumstances is essential. In particular, information may gain an entirely different sense and quality depending on the situation and the recipients who can have less or additional knowledge - not to mention the fact, whether the concerned subject wants this piece of information about himself revealed to these recipients at all (Bundesverfassungsgericht, 1980).

"Every individual should in principle – without restriction to its private sphere – decide on its own how it wants to present itself to others or the public and to what extend others have his personality at their disposal" (Bundesverfassungsgericht, 1980).

In 1970, the world's first law concerning the protection of personal data was enacted in the state of Hessen, Germany. The first federal law dates from 1977 (Kahlert, 2004). In 1983, the Federal Constitutional Court derived the right for protection of personal data from the constitution (Grundgesetz, GG), more precisely from the right of free development of personality (Art. 2 GG) and from the human dignity (Art. 1 GG) (Tinnefeld, 2005).

Especially nowadays when by automatic data processing loads of data are generated which can effectively be combined to personal profiles and with the affected person having no chance to control it for correctness and (mis)use (Kahlert, 2004) the requirements for protection of personal data get to some new extent.

2.2 The Legal Situation in Germany

The entire subject of data privacy regarding the retainment of IP address traces is entirely disputed and there are no judgments of higher courts creating certainty for the practice. It begins with the question whether IP addresses are personal data at all, continues with the controversially discussion about which laws are corresponding to the transmission on the IP layer, and ends with the problem which legal admissions are applicable under which prerequisite. Therefore a complex and thorough analysis of the relevant legal regulations and opinions on these subjects is necessary. Our study provides an overview to the current state – in this aspect, we discuss our own conclusions.

2.2.1 Relevant Regulations

In Germany, many different laws are involved: the EU Directives 95/46/EG and 2002/58/EG, the Federal Data Protection Law (Bundesdatenschutzgesetz, BDSG), and the laws of the federal states (Landesdatenschutzgesetz, LDSG), e.g. of Baden-Württemberg (LDSGbw). Besides those general laws, there exist special laws like the Telecommunications Act (Telekommunikationsgesetz, TKG) and the Teleservices Act / Teleservices Privacy Act (Teledienstegesetz / Tledienstedatenschutzgesetz, TDG/TDDSG) as well as the Treaty on Media Services (Mediendienste-Staatsvertrag, MDStV). Finally the constitution serves as a guideline for the application of these laws.

The capability to relate data about personal or matter-of-fact relations of a natural person to this person is the connection factor and prerequisite for the applicability of any privacy related regulation. Only if excessive expenditure is necessary to determine the concerned person, data is considered anonymous (cf. § 3 VI BDSG). If the concerned natural person cannot or can only with disproportional large efforts be determined, their privacy cannot be interfered with.

2.2.2 The IP Address as Personal Data

For packet traces, the IP address is the piece of data in question. How far is it suitable to create a link to a natural person? It has to be noted that the possibilities of the processing unit are the criterion for this decision (Dammann, 2003). The same data may be related to a natural person by one unit but not by another. E.g. in criminal processes, the state can demand log files and account data from different ISPs and track the user by combining this data, while private organizations do not have these legal authorities. In the case of transmission of information, the receiver's possibilities to relate the data to a natural person are authoritative for the question if it is considered as personal data (Gola, 1997).

Considering IP addresses belonging to the data processing entity's domain, this poses no problems. The entity assigns the IP addresses, has access to both, log files and account data. Evidently, the self-administered IP addresses and their related data are linkable to a certain person. For other, e.g. external, IP addresses the situation is more complicated and controversial. In this case, many distinguish between dynamical and static IP addresses.

Because the dynamic IP address changes each time a host connects to the Internet, it is more difficult to establish a user profile or to collect data of the same host because it cannot be re-identified. Therefore, most opinions decide dynamic IP addresses not capable to create a link to the user for others than the Internet access provider. They would

need the provider's assistance (Hoeren, 2001), who however is not allowed to transmit personal data to them if there are no legal permissions.

Based on the captured IP address information, one can only find out which Internet access provider the user belongs to. If the provider only serves a specific region, the user is supposed to be located there (Schläger, 1998). But this is far from identifying the user and infringing his rights. For example, without additional data gathered from login forms people running a web server cannot identify the user solely based on his IP address. Therefore, many say that the use of dynamic IP addresses make an anonymous or at least pseudonymous use of Internet services possible (Tinnefeld, 2005). In contrary, some others believe that access providers would help in revealing the user's identity (Schaar, 2002). Nevertheless, such help would act contrary to the law.

Accidentally revealing user data is unlikely because people are already sensitized to this issue. Some talk about the "many and various possibilities of joining data in the Internet" (Arbeitskreis Medien, 2002) without giving actual reasons how the identification could be feasible.

Static IP addresses are considered as capable of identifying the user by the majority of opinions. Some do not give any reasons or simply call it obvious. Others come up with the argument that data about a user can be easily collected over a longer time period because the address remains the same (Roßnagel, 2003). This in turn should permit his identification. Therefore, logging of static IP addresses would be inadmissible because of privacy laws (Unabhängiges Landeszentrum für Datenschutz Schleswig Holstein, 2000). This, finally, leads to the opinion that any logging of IP addresses would be against the law because dynamic and static addresses cannot be distinguished (Bahr, 2003).

But none of these opinions really state any methods by which the user's identity can be revealed. The long-term static IP addresses do not help that much because every storage unit can only obtain information from direct interactions with associated servers. And because the address is not known to be static, it is also unknown that it is the same user that visits the server again. Even the fact that the user has the same address on other sites does not help them to identify him without further information.

Finally, an IP address is only related to a host but not to a user. Dynamic addresses change and static addresses normally cannot be identified. The hosts might as well be in computer pools in universities and libraries or publicly accessible like in internet cafes. Private computers are partly used by several persons. In many access networks, network address translation is employed so many persons use the same IP address in the internet. Even in the case of single user with a static address it is obscure how he can be discovered without any additional information. Because

the address is not publicly known as static, one does not know if the data belongs to different persons or only to a single but complex personality.

2.2.3 Conclusions

Which possibilities exist to identify a user of a given IP address? The domain and the host name can be determined using the DNS (domain name system). Often, the domain describes the Internet access provider. A whois query allows finding the owner of the domain. But normally there are many hosts or users inside this domain and the server of this domain is rarely used by a single user.

The user can be re-identified by the use of cookies. But these cookies only allow identifying the user at further visits of collecting unit's server. But without additional data from other sources (e.g. login forms) this cannot help to find the real-life person. In the case of web servers the referrer URL is a hint to another source of data. But the institution running the other server is not allowed to give away the user's data by the law.

Against common opinion for external IP addresses no effective way is found to discover the person using a given IP address. Therefore, external IP addresses in log files are not subject to privacy regulations if the processing unit has no additional data to create a link to the users.

2.2.4 Legal Admissions and Modalities for Processing Personal Data

For the IP addresses, which can be related to a natural person (e.g. local addresses) and therefore form personal data, a legal admission or the user's consent is necessary for collecting, processing, and storing such information. The need for permission is the basic principle for collecting and processing personal data in both, the EU directives and the national law (Tinnefeld, 2005). Because it is impractical to obtain each user's consent for publicly available services one has to rely on legal admissions.

If IP addresses on the network layer of the ISO-OSI model are regulated by the TKG (telecommunications act) or by the TDG/TDDSG (teleservices act/teleservices privacy act) as well as the MDStV (treaty on media services) is discussed very controversially (for details refer to Haibl, 2005). On principle, the TKG deals with the technical process of the telecommunication while the TDG/TDDSG and the MDStV concern the teleservice (Brunner, 1999). Because IP numbers are used as addresses for the transport of data through the Internet disregarding the transmitted content and do not provide higher layer services they rather should be comprised by the TKG's regulations.

Layer	Example	Applicable Law
content layer	form data	BDSG, LDSG
service layer	click streams	TDG/TDDSG, MDStV
network layer	telecommunication as transmission service	TKG

Table 1: data privacy layer model (Federrath, 2004)

In the TKG, there exist several admissions for collecting, processing, and storing data. In the field of accounting, there exists only one admission pertaining to billing (§ 97 TKG) which is not suitable for telecommunication services that are free of charge. § 100 TKG permits also the collection of data in case of obtainment of service by fraud if there are concrete indications. Concerning intrusion detection, § 100 TKG can be considered as an admission for the processing of personal data regarding failure detection and treatment. Unfortunately, this is generally disputed. § 28 BDSG (Federal Data Protection Act) permits the processing of personal data if it is necessary for the legitimate interests of the processing unit and the interests of the concerned person are not violated. These legitimate interests can originate from § 9 BDSG and § 109 TKG which oblige the processing unit to take measures to protect personal data and the telecommunication facilities. In nearly any computer system or local network, there exist personal data which has to be protected. It is contested whether the general permission in § 28 BDSG is applicable or if it is superseded by the more specific regulations in the TKG. However, one can differentiate (Schaar, 1999): The admissions in the TKG apply only to personal data in the relationship between user and service provider but outside of this relation, e.g. in case of attacks or other security issues, the general rule should remain applicable. In addition, personal data may be collected and processed in order to enable research if it is necessary for the research and if the interest in the research outweighs the interests of the concerned person (§§ 28 BDSG, 35 LDSG). In this case, the data has to be anonymized as soon as possible (§§ 40 BDSG, 35 LDSG).

For the processing of personal data, basic principles are stated in § 3a BDSG: the principle of avoidance and thriftiness. The resulting modalities for the processing of personal data include the necessity that the collected data has to be pseudonymized, anonymized, or deleted as soon as possible if the intended purpose can still be accomplished. For many purposes like traffic analysis, only the communication pattern, especially the subnets, is of interest. The plain IP addresses are not necessary and pseudonymization and anonymization must take place.

2.2.5 Outlook / Future

In future, the privacy of communication is to be cut down by the European Union's directive 2006/24/EC on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks. With these plans, the divergence between politics, law, and technical feasibility becomes noticeable. Most installed networks are not suitable for logging of the IP headers and would require substantial chances. The costs for this plan pose serious problems, too. For example, the COLT Telecom Group estimates expenses of more than 90 million Euro per year for the retainment of the IP addresses not including costs for additional systems or the retainment of data of its switched connections (E-Commerce-Magazin, 2005). Considering a turn over of 1970 million Euro of the Colt-Group in 2004 (COLT, 2005), the financial impact of the EU's plans comes clear and their feasibility is questionable.

With the transposition of the directive to national law by the member states, the legitimation of retention by the affected ISPs is out of question. But the admissions discussed above are still necessary for the processing of the data regarding its concrete purpose which has to be admitted separately from the mere retention.

3. Anonymization of Packet Traces and Measurement Results

There are different techniques that can be thought of to anonymize the IP addresses: the truncation of IP addresses, the storage of corresponding hash values, encryption, or the mapping by an injective and prefix preserving function.

The truncation of the address is simple and efficient. By sufficient shortening the address a group of users is mapped to the same imperfect address. Thus, the single user cannot be identified anymore. The advantages of this proceeding are that the network prefixes are preserved for traffic analysis and the data remains simple for further interpretation. The anonymized data is available to any use. If some addresses are conspicuous, the full address can be recorded in accordance to privacy laws.

Another idea is to calculate and store the hash value of an address, e.g. using cryptographic hash functions such as MD5. This has to be done separately for source and destination address in order to make end points comparable and not only the connection. Otherwise connection patterns cannot be discovered. But it is questionable if anonymity can be reached by hashing because the input range for the (well-known) hashing algorithm is bounded. A good hash function will evenly distribute these input values to an output range larger than the input range. So a sort of injective mapping occurs. On a standard Pentium IV, the MD5 hash values for all 4 billion IPv4 addresses can be calculated in less than 2 hours. So it is easy to build a lookup table for the original addresses. Therefore, no anonymity is achieved

by hashing. This rule also applies to IPv6 addresses: even if it becomes hard to create a full lookup table, in most scenarios smaller address ranges can be distinguished, e.g. the address space of a university. Narrowing the function's output range produces collisions which is similar to truncating the addresses. Nevertheless, this is more complicated than simple truncation and comes up with no advantages. Network prefixes are not kept.

Key-based encryption of the IP addresses might handle the problem that a lookup table can easily be created. But as long as the key is kept for further encryption in order to achieve comparability with the already encrypted addresses, a lookup table can still be built. After a key change the old and the newly encrypted data are not comparable similarly to IP addresses from different sites using different keys. Another disadvantage of encryption is that the prefixes of the addresses are lost. Moreover, it is highly questionable if anonymity is achieved. The encryption creates a one-to-one mapping to the addresses so statistical attacks based on users customs (time, destination addresses) are possible. Without knowledge of the key, a user's ciphered IP is identified and the data can be filtered for its connections to the possibly unencrypted external addresses. One-way functions and key-based one-way functions suffer from the same problems as hashing and encryption.

Another approach suitable for the analysis of communication patterns is to use a key based prefix preserving mapping of the IP addresses (Xu, 2002). While communication patterns are kept, this method suffers from the same problems as encryption: the lack of comparability after a key change or of data of different sites. The work's goal was to make the publication of the traces possible. Because the used keys are unknown to others the data may be freely distributed for analysis, other sites will not be able to recognize even their own addresses.

The anonymization process has to be fast in order to permit anonymization of data in real time. Therefore, it is possible for any key based algorithm while the original key is kept to build a lookup table. The IP address which is a pseudonym would only be replaced by another pseudonym. No additional privacy is achieved. Key-based approaches lack comparability if different keys are used or keys are changed. Any injective mapping of the IP addresses is vulnerable to statistic analyses for the collecting unit if the behavior patterns of the users are known.

For this reason an algorithm creating an uncertainty is favorable like the truncation of the address if no injective mapping is necessary. The truncated addresses are easy to interpret and the shortening can varied according to privacy needs. It is fast and keeps even the original network prefixes.

4. Discussion

4.1 Pragmatic solutions

As a practical advice, the following procedures can be followed. The purpose of the intended processing of data should be stated and evaluated to what extend personal data is required to achieve this purpose. For example, for measurement and traffic engineering only the traffic between sub-networks is of interest but not the single IP address. So the needs for storing and publishing personal data can be limited. For other purposes, aggregation of the data can eliminate its personal quality. For all remaining personal data, the national law has to be checked for admissions. If the processing of personal data is allowed, its modalities have to be oriented on the principle of avoidance and thriftiness. Finally, the processing has to be secured by technical and organizational measures. Access can be limited to some authorized persons and to the data they need. Technical measures are mechanical barriers or password mechanisms. Finally, the personal data has to be deleted if it is no longer necessary, e.g. in the case of intrusion detection older data is of no need if no occurrences took place. Perhaps (legal) maximum periods for the retainment have to be obeyed.

4.2 Need for unified anonymization techniques

Packet traces from different locations in the network are needed for effective analysis and to gain a ‘global view’ of the network. The data in turn needs a uniform format for easy interpretation and efficient processing of data from different sources. These traces need to be anonymized to make them online available without risking to offend the law. The solutions are unified anonymization techniques employed by many processing units so a large amount of traces would be available for research and can be used without much effort.

4.3 libAnon, a first step towards anonymized and comparable data

With the libAnon (Haibl, 2005), a versatile framework for the application of different anonymization algorithms was created. The library allows to parameterize and to apply various anonymization algorithms according to source and destination IP addresses and ports. The degree and manner of anonymization can be chosen for the special needs of the processing unit. Data from sparsely used IP address ranges and not homogenously used networks may need stronger anonymization than data from large or dense and homogeneous networks. The port information may be filtered out, too, if the connections to certain ports are likely to reveal the users identity.

The library supports the retainment of a variable length network prefixes and implements the algorithms for truncation, encryption by blowfish, and (for testing purposes) MD5 hashing and a simple remapping.

5. Conclusion

In this article, we discussed the needs for anonymization techniques in the domain of network measurement and monitoring. Our study is based on a review of the current legal situation in Germany concerning privacy regulations. A historical overview to the appropriate privacy laws shows that regulations are very strict if a specific person can be assigned to stored monitoring information.

Generally, many effective network monitoring and measurement methods and tools have been proposed in the last decade demanding for distributed analysis of measurement results and packet traces. To achieve this goal, network data from different sources is needed without inflicting individual rights of data privacy. This can be accomplished by data anonymization that permits a large-scale analysis of network data. Our study reveals necessary anonymization and pseudonymization techniques that correspond with the current privacy regulations.

In conclusion, it can be said that well-structured and standardized anonymization mechanisms are needed (1) for observing the law in terms of privacy; and (2) for allowing a publication of packet traces that is still comprehensive and comparable.

References

Anagnostakis, K. G., S. Ioannidis, et al. (2002). Efficient Packet Monitoring for Network Management. 8th IEEE Network Operations and Management Symposium (NOMS), Florence, Italy.

Arbeitskreis Medien (2002), "Orientierungshilfe zum Umgang mit personenbezogenen Daten bei Internetdiensten", in Zezschwitz, F. v. (Ed.), "Einunddreißigster Tätigkeitsbericht des Hessischen Datenschutzbeauftragten", Wiesbaden. Available http://www.datenschutz.hessen.de/Tb31/K25P03.htm

Bahr, M. (2003), "IP-Speicherung durch Webseitenbetreiber rechtlich zulässig?", Hamburg. Available http://www.Dr-Bahr.com/download/ip_logging.pdf

Brandeis, L.D. and Warren, S. (1890), "The Right to Privacy. The Implicit Made Explicit", Harvard L.R., Vol. IV No. 5. Available http://www.lawrence.edu/fast/boardmaw/Privacy_brand_warr2.html

Brunner, S. (1999), E § 2, in Manssen, G., (Ed.), "Telekommunikations- und Multimediarecht", Schmidt, Berlin.

Bundesverfassungsgericht (1980), Neue Juristische Wochenschrift, pp. 2070 et seqq.

COLT (2005), “COLT Annual Results – Annual Results 2004”, COLT Group, London. Available http://www.colt.net/investor_relations/annual_results

Damann, U. (2003), § 3, in Spiros, S. (Ed.), “Kommentar zum Bundesdatenschutzgesetz”, Nomos-Verlags-Gesellschaft, Baden-Baden.

E-Commerce-Magazin (2005), “Fast jeder zweite Bundesbürger lehnt EU-Pläne zur Vorratsdatenspeicherung ab”, WIN-Verlag GmbH & CoKG, Vaterstetten. Available http://www.e-commerce-magazin.de/index.php3?page=news-show_neu.php3&naechster=8495

Federrath, H. (2004), “Gesetzliche Grundlagen”, University of Regensburg. Available http://www-sec.uni-regensburg.de/security/Folien/02Recht.pdf

Gola, P. and Schomerus, R. (1997), “Bundesdatenschutzgesetz”, 6th ed., Beck, München.

Haibl, F. (2005), “Studienarbeit Anonymisierung – Erstellung einer Funktion zur regelbasierten Anonymisierung von Verbindungsdaten”, Wilhelm-Schickard-Institute for Computer Science, University of Tübingen.

Hoeren, T. (2001), ”Grundzüge des Internetrechts: E-Commerce, Domains, Urheberrecht”, Beck, München.

Kahlert, H. (2004), “Als Big Brother noch keine Fernsehsendung war“, Forum Recht, Vol. 3, pp. 76-78, Hamburg

Pang, R. and V. Paxson (2003). A High-level Programming Environment for Packet Trace Anonymization and Transformation. SIGCOMM 2003, Karlsruhe, Germany, ACM Press.

Paxson, V., J. Mahdavi, et al. (1998). "An Architecture for Large-Scale Internet Measurement." IEEE Communications Magazine 31(8).

Roßnagel, A. (2003), “Handbuch Datenschutzrecht”, Beck, München.

Schaar, P. (1999), § 4 TDDSG, in Roßnagel, A. (Ed.), “Recht der Multimediadienste”, Beck, München

Schaar, P. (2002), “Datenschutz im Internet”, Beck, München.

Schaar, P. (2004), “BfD-Info - Bundesdatenschutzgesetz - Text und Erläuterung“, 11th ed., Bundesbeauftragter für Datenschutz, Bonn.

Schläger, U. (1998), “Datenschutz im Internet”, Fachberatung Datenschutz und Informationssicherheit, Bremen. Available http://www.tib-hamburg.de/datenschutz1.htm

Schmitz, P. (1999), 16.4, in Hoeren, T. (Ed.), “Handbuch Multimedia-Recht”, Beck, München.

Tinnefeld, M.-T., Ehmann, E. and Gerling, R. (2005), “Einführung in das Datenschutzrecht – Datenschutz und Informationsfreiheit in europäischer Sicht”, 4th ed., Oldenbourg, München.

Unabhängiges Landeszentrum für Datenschutz Schleswig Holstein (2000), "22. Tätigkeitsbericht", Kiel. Available http://www.datenschutzzentrum.de/material/tb/tb22/kap7.htm

Xu, J., Fan, J. Ammar, M., and Moon, S. 2002. "Prefix-Preserving IP Address Anonymization: Measurement-based Security Evaluation and a New Cryptography-based Scheme", Proc. of 10th IEEE International Conference on Network Protocols (ICNP 2002), Paris, France. Available http://www.cc.gatech.edu/~jx/reprints/ICNP02A.pdf