IDENTIFICATION IN CYBERSPACE

by

ALŽBĚTA KRAUSOVÁ*

Cyberspace could be compared with a fancy-dress ball. Before entering this different world everybody chooses a mask of some appearance, a certain social role and comes in with a new identity. As there are many kinds of people, we can find virtual identities which truly correspond to real persons in almost every aspect. However, majority of people use illusory anonymity of the cyberspace to enjoy the freedom to be anybody they want to be. Such people create various masks painted with false colours. These people try to avoid revelation of their real identities. Ways of their identification by means of law are not sufficient and effective yet.

In this paper the author is going to focus on non-legal means of identification, namely on technical means and on identity detection with help of publicly available means. Legal conditions of using these means shall be also clarified. The author will try to explore informative value of data acquired this way and answer the question if they can be used as a body of evidence at court.

KEYWORDS

Disclosure of information, identification, identifiers, personal data, private investigative techniques, pseudonym

* Mgr. Alžběta Krausová, a researcher working for ICRI – K. U. Leuven – IBBT. The paper was reviewed by Prof. Dr. Jos Dumortier, the director of ICRI (Interdisciplinary Centre for Law and ICT), a research centre within the Faculty of Law, Katholieke Universiteit Leuven. Contact: Betty.Krausova@law.kuleuven.be

THE FREEDOM TO BE ANYBODY AND OPERATION OF LAW [1]

We live in an amazing world. Technical means allow us to communicate almost with anybody on this planet, to share information, to learn about the world, generally said to achieve our goals in society much easier. The "magical" phenomenon called Cyberspace provides us with the freedom of speech, the freedom to be anybody we want to be which also includes the freedom to be ourselves just as we are. From a point of view of an average person this may sound great, from a lawyer's point of view unfortunately this sounds rather dangerous. Why dangerous? Because the freedom to be anybody we want to be breaks the general principle of operation of law.

The freedom to be anybody does not only include the freedom to be ourselves just as we are but a possibility to change virtually our own identities as well. This change is usually done the way that a particular individual chooses a pseudonym and identifies herself with this pseudonym in the cyberspace. The connection between herself and the pseudonym creates the particular individual person on her own and it depends on her whether she will disclose her personal data to the world or to a relevant internet service provider (ISP) and / or whether she will give accurate information about herself. She can make herself to be absolutely unrecognizable in the virtual environment. However, when a natural person cannot be associated with her manifestations, this person is in the scope of her virtual manifestations excluded from the operation of law, and thus she can violate rights of other people at no punishment.

For a better understanding the problem is going to be presented on a case study. Let's say there is a girl called Ine (In-visibl-e) who has just created her new identity (i. e. some new user account) in the privacy of her room when using her own computer. As she has not revealed any personal information in the new identity's profile she considers herself to be unidentifiable. She does not presume that anybody would be able to associate her manifestations in the virtual environment with her as a natural person in the offline world.

Let's imagine a situation when Ine knows some person called Victoria (Vict-im). On one beautiful day Ine starts to hate Victoria. Ine decides to make Victoria's life very unpleasant and she starts to write her emails of

such a nature that these emails violate Victoria's personal rights (e. g. right of integrity, right of privacy). To prevent legal consequences of her wrongful acts Ine sends these emails via her email address that she created without revealing any personal information with such a pseudonym that it is impossible to connect this email address with Ine.

Let's turn to Victoria's situation now. Her personal rights are violated and she has no civil legal remedy as she cannot start legal proceedings against someone hidden under a nickname Secretdaemon666@msn.com. Or can she?[1]

This is a crucial question that is related to the aim of this paper: to find a model of searching and a way of connecting individual identifiers of an unknown person so this person can be identified and made reliable for her acts in the cyberspace at court. This model should help to enforce individuals' rights, to claim indemnities, to prevent abuse of the freedom to be anybody and violation of other people's rights.

JUDICIAL PROTECTION OF RIGHTS [2]

Modern law provides means of protection for individuals who are being harmed. However, in accordance with the principle "*vigilantibus iura scripta sunt*" law imposes requirements on people who apply for protection.

In European countries laws usually state that a subject may start civil judicial proceedings only against a precisely identified defendant. For Victoria it means that in case she wants to start legal proceedings against Secretdaemon666@msn.com, she must at first find out, who that Secretdaemon666 (hereinafter only SD666) is.[2] For example, pursuant to the Czech, Slovak, Belgian or Italian law, in the petition she must state a name, a surname and a place of residence of the defendant.[3]

---

[1] In the following text only the situation of civil infringement of personal rights is analyzed. Depending on contents of a harmful email an individual may seek protection through provisions of criminal law. For example defamation is in various laws regulated as an infraction (§ 49 of the Czech act no. 200/1990 of the Coll., on Administrative Infractions as amended, §3 of the Upper-Austrian Police Criminal Act) or even considered to be a crime (a crime of "Ingiuria" in accordance with an article 594 of the Italian Criminal Code).

[2] In some countries where there exists a legal institute of a tort, a help of state bodies is provided to citizens so they do not need to identify the tort-feasor by themselves.

[3] See § 79 paragraph 1 of the Czech act no. 99/1963 of the Coll., Civil Procedure Code as amended, § 79 paragraph 1 of the Slovak act no. 99/1963 of the Coll., Civil Procedure Code as amended, an article 43 of the Belgian Civil Procedure Code or an article 163 of the Italian Civil Procedure Code.

At this moment Victoria has to identify the defendant. In this paper the process of identification is understood as associating a virtual manifestation of some entity to a real "offline" entity which can be found and made reliable for her acts performed in the cyberspace.

However, a person like Victoria does not have any chance how to identify SD666 with help of a legal means. With the term "legal means" a legally regulated way of identification and authentication of a sender is understood (namely an advanced electronic signature).

Legal means of identification is used to prove a connection of some natural person with a virtual identifier. Its usage is necessary when somebody acts in the cyberspace and intends these acts to be legally binding. As an example a transfer of money can be mentioned. Advanced or certified electronic signatures are used between parties who need to know the other entity's identity to possibly enforce the terms of a contract.

The presented case is though completely different, as in this case Ine tries to prevent revelation of her true "offline" identity. Thus, the first question is whether there is any chance for Victoria to find out who that SD666 is. For a successful identification Victoria has to use non-legal means. For purposes of this paper the term "non-legal means" is defined as performance of any process that leads to acquiring some information about an unknown person, in other words the term "non-legal means" refers to private investigative techniques.

In order to protect her rights Victoria has to use non-legal means of identification and persuade a court that her way of identification, the way how she linked data received from the cyberspace is correct and that a defendant stated in her petition is the one who shall be made reliable for the manifestations made in the virtual environment.

THE PROCESS OF IDENTIFICATION [3]

When identifying a person one should at first focus on the very process of identification to find out whether the identification in a particular case is even possible. As it was mentioned, the process of identification in this paper is defined as associating a virtual manifestation of some entity to a real "offline" entity. Speaking in a natural language, it means establishing that

some emails or posts on forums or other discourses were made by a particular natural person.

Process of identification has its own specifics in various environments. In the environment of the cyberspace one can start from a presumption that everybody acting in the cyberspace possesses some set of identifiers. Good examples of these identifiers are accounts for various internet services, namely one or more email accounts (usually an official and an unofficial email account), nicknames on forums, chat rooms, blog accounts, instant messaging accounts (ICQ, MSN, Skype, etc.) and so on. The mentioned identifiers have one aspect in common: they are unique for each person. They are characterized by exclusivity and rivalry. However, a single identifier explored separately is usually not sufficient for identification of a natural person.

For more reliable identification one need to acquire more information about an unknown person preferably by means of collecting various identifiers that provide additional user-related information. The most reliable way of identification is to exploit a phenomenon called by technical experts on privacy as "an accidental identifier collision".[4] In a natural language one can say that it is a situation when identifiers related to a person are not independent but instead they (or at least some of them) are interconnected. The means of the interconnection is then some disclosed user-related information.

SUCCESS AND FAILURE FACTORS [3.1]

As in any other process in the process of identification there also exist success and failure factors.

The key factor for the successful identification lies in disclosure of data by a person who tries to hide herself. In communication certain disclosure of data (or better say unintended leakage) happens every time although sometimes it could seem that no data was disclosed. In virtual communication there are more sources of acquiring data for identification. These sources can be divided to two categories: identifiers themselves and user-related information. User-related information can be divided into the follow-

[4] Brands, S. 2006, Secure User Identification Without Privacy Erosion, *University of Ottawa Law & Technology Journal*, vol. 3, no. 1, p. 211, Available at: http://ssrn.com/abstract=999695

ing subcategories: technical data connected to a message created by a person using a particular identifier, information about a person disclosed in a profile related to an identifier (e.g. web pages related to a unique ICQ number), information acquired from search engines linkable to an identifier, and the very contents of the message. In the contents an author does not normally disclose any clear identifier such as her name, a phone number or a date of birth and so on but the contents indicate aims, attitudes and feelings of the author. That can give some guidance when determining the true identity of the author.

The next factor increasing a chance of the successful identification is a possibility to link data acquired from different sources. This corresponds to the accidental identifiers collision mentioned above.

Traceability of disclosed data is another factor of the successful identification. With the term "traceability" an identification of "the IP address that caused an action to occur"[5] is understood. In other words it means a possibility to find the IP address on which behalf a message was sent. However, one must be aware of the fact that not every kind of data can be traced. Traceability of disclosed data thought increases a chance to reveal the true identity of a sender.

The process of identification has significant barriers though. Advanced technical skills and knowledge of a sender represent the first barrier. If a sender is aware of privacy issues, risks and principles of anonymous behaviour in the cyberspace and if she has technical knowledge as well she can use e.g. anonymous remailers or connections through proxy servers so her acts cannot be tracked. She can prevent the identifiers collision as well so she will stay hidden. The next barrier is recipient's lack of technical skills and knowledge how to search the cyberspace. The point is that even if a sender disclosed enough information to be identified by a recipient, the recipient is not able to use techniques that would lead to identification of the sender.

[5] Clayton, R. 2001, The Limits of Traceability, Available at: http://www.cl.cam.ac.uk/~rnc1/The_Limits_of_Traceability.html

THE MODEL OF IDENTIFICATION [3.2]

After being aware of both success and failure factors one can focus on the very model of identification. The model presented in this paper has three phases. First two phases can be parallel as information acquired in the second phase can retroactively influence the first phase.

The first phase begins at the moment of receiving a harmful email. The primary task is to analyze contents of the message and to try to determine possible authors. The question is: "*Cui prodest?*" In other words one must ask who has a motive to write such a message to a receiver. The outcome of this phase shall be in a form of a list with names of potential authors. At every person's name all available information concerning personal identifiers or personal data should be mentioned (the most important information is of course an address as it is needed to start legal proceedings, but every piece of information will later help when associating profiles).

In the second phase one should create a profile of the author. When creating such a profile it is necessary to collect identifiers and user-related information. One can do that with help of the private investigative techniques. The portrait of the author should ideally consist of identifiers like an official email address, a phone number, and a residency address. Usually it consists of various email addresses, instant messaging identifiers, and nicknames, information about hobbies, interests, and cyberspace friends of the author and so on. These identifiers and personal information may seem unsubstantial at the first sight. However, every piece of information can be helpful and lead to acquiring a more detailed profile.

At the third phase the profile created in the second phase is associated with some person from the list created in the first phase. Association of course should be based on similarities in both profiles. The most reliable method of associating is to compare information in both profiles with a same denomination (a name, an address, a phone number, etc.) and if the values are consistent, the connection is created. Sometimes a value is not precise, but only approximate, or partial (an address of a person from a list is full but information about an address of a person hidden behind the profile is only a name of a city). During the process of associating it usually becomes clear very soon which people should be excluded as almost none of

their values are consistent with values of the created profile so the connection between the profiles is weak. Associating is the more successful the more values are corresponding.

THE PRIVATE INVESTIGATIVE TECHNIQUES [3.3]

In the previous text the "private investigative techniques" are mentioned. Their purpose is to help with creating a profile of the author of an email infringing personal rights. This subsection will describe the said techniques more in detail. These techniques can be basically divided into two categories: technical means and publicly available means.

For the purposes of this paper the term "technical means" is defined as performance of technical procedures leading to acquiring necessary information on the basis of knowledge of a cyberspace architecture and its principles. For using of this means one usually needs to be educated in the field of information technology. That is why this means constitutes a separate category. The most useful technical means in the presented case study is a technique of tracking emails. As servers that provide this service to individuals who cannot track emails by themselves[6] already exist, this service becomes available for more people. When using this technique one can find out from which town the email was sent and who is the ISP providing an internet connection for the computer from which the message was sent.

The other category of private investigative techniques can be called as searching with help of "publicly available means". This term shall express that the cyberspace can be perceived as a huge database full of publicly available information disclosed by ordinary people, all types of organizations, as well as administrative bodies and state institutions. Anybody can search this database in the parts that are publicly available. To search the cyberspace no special education is needed, anybody can use common search engines[7] and get information without any special effort. Except the search engines also special public databases created by state authorities should be mentioned. These online available databases (Commercial Register, Land Register, Register of Experts and Interpreters, etc.) contain reli-

[6] See for example URL *www.whatismyipaddress.com*. This server provides the service of tracking emails for free. All you need is to provide a header of the tracked email.

[7] The already exists a new verb for searching information: „to google" something.

able information as they were created by a trustful third party. However, the problem is that in many cases it is possible to search these databases only when one already possesses some identifier concerning the person whose true identity shall be determined and at the same time this identifier is of the consistent denomination as some of searching criterions in a particular database.[8]

When using private investigative techniques and creating a profile of some person one must be very careful as involvement of wrong information can occur. This can happen when two or more people possess some same identifier. With the same identifier it is especially meant that they have the same combination of their first name and surname (for example the most common surname in America is "Smith" and the most common male first name is "James"[9]). Another mistake can be made when one presumes that identifiers with the same wording but provided by different ISPs belong to the same person (e.g. without any additional evidence one cannot presume that a controller of an identifier SD666@msn.com is the same person as a controller of an identifier SD666@gmail.com). In this phase all information must be very carefully evaluated. Criteria for evaluation are a source of information and a context.

Concerning the source it is useful to apply a basic principle that information disclosed by a third party is more reliable than information disclosed by a person that wants to hide herself. However, every piece of information must be evaluated in relation to another piece of information and in relation to the complex view on an emerging profile that should be consistent. In other words, the context must be taken in account.

LEGAL ASPECTS OF THE PRIVATE IDENTIFICATION PROCESS [4]

This section will focus on legal regulation of the private investigative process leading to determining someone's identity. The main purpose is to find out whether there are any limitations for the person trying to defend her own rights. The aim of this section is to explore whether such acting is regu-

[8] Concerning searching in special public databases a situation in the Czech Republic is described. In another countries situation may differ.

[9] This information is provided by U.S. Census Bureau, Population Division as of the year 1990, retrieved from http://www.census.gov/genealogy/names/names_files.html

lated by the Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Applicability of the Directive 95/46/EC is determined in its Article 3. The first paragraph defines scope of the Directive positively: "This Directive shall apply to the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data, which form part of a filing system or are intended to form part of a filing system."

The definition of personal data in accordance with this Directive is following: "personal data shall mean any information relating to an identified or identifiable natural person"[10] whereas "an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[11]

When determining whether the Directive is applicable on the private identification process the term "identifiable person" is very important. In our case study Victoria, who is looking for the true identity of Ine, does not know that the data she is collecting is related to Ine. For Victoria this data is only a means to identify someone. It becomes personal data at the moment when Victoria reliably identifies Ine. The following court proceedings later only declare whether the identification performed by Victoria was correct or not.

As the applicability of the Directive is excluded there are no obstacles in performance of the private identification process. Victoria then can finish the process of identification and finally commence court proceedings and exploit her right to judicial protection.

COURT PROCEEDINGS [5]

As it was already mentioned, for initiating court proceedings a defendant's name and an address of residency must be stated in a petition. Let's presume that Victoria used the proposed model of identification and that she was successful because Ine had disclosed some other information about her-

[10] Article 2 of the Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[11] See the previous footnote.

self and an accidental identifier collision had occurred. Victoria brings an action against Ine. In the petition she states Ine's name and address, describes facts of a case, and in a prayer of the petition she claims indemnities.

When the petition is delivered to a court,[12] the court examines whether legal requirements are fulfilled. If Victoria would bring an action against SD666, the court would not proceed with a trial due to a defective petition. At this moment the court has to execute proposed evidences and evaluate them.

The importance of the identification process lies in the fact that on its basis court proceedings can be at least started. A plaintiff can propose the court to execute special pieces of evidence to find out information which she could not acquire without help of court. A court is for example entitled to ask an ISP to provide information about a name and a residency address of person using particular IP address.

The success in such trial of course depends on precise and persuading argumentation. The chance to succeed is then more or less on the same level as in a common case that does not concern determining of online identities.

CONCLUSION [6]

From the above presented case study it is quite obvious that a person whose civil rights were or have been violated by someone else in the environment of the cyberspace and who is not protected by the legal institute of a tort has to make considerable efforts to seek a remedy. In the author's opinion law imposes undue requirements on people injured this way. As law is silent concerning disputes arisen in the cyberspace, people have to find their own way how to catch an infringer. The situation for people harmed this way would be much easier if law would set special rules concerning disputes arisen in the cyberspace.

The easiest way is to set these rules in a national Civil Procedure Code. As the courts are known for their resentment to hear a case concerning online activities, the author would recommend setting a duty of non-refusing or duty to proceed a trial. The courts should be obliged to hear both parties and should have a power to propose and examine own evidence.

[12] The following procedure is described in accordance with the Czech procedural law.

REFERENCES

[1] Act no. 140/1961 of the Coll. (Czech Republic), Criminal Code as amended

[2] Act no. 99/1963 of the Coll. (Czech Republic), Civil Procedure Code as amended

[3] Act no. 99/1963 of the Coll. (Slovak Republic), Civil Procedure Code as amended

[4] Act no. 200/1990 of the Coll. (Czech Republic), on Administrative Infractions as amended

[5] Act no. 101/2000 of the Coll. (Czech Republic), on the Protection of Personal Data as amended

[6] Brands, S. 2006, Secure User Identification Without Privacy Erosion, *University of Ottawa Law & Technology Journal*, vol. 3, no. 1, pp 205-223, Available at: http://ssrn.com/abstract=999695

[7] Clayton, R. 2005, *Anonymity and traceability in cyberspace*, Technical report no. 653, University of Cambridge, Computer Laboratory

[8] Clayton, R. 2001, *The Limits of Traceability*, Available at: http://www.cl.cam.ac.uk/~rnc1/The_Limits_of_Traceability.html

[9] Directive 95/46/EC of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data

[10] Directive 1999/93/ECof 13 December 1999 on a Community framework for electronic signatures

[11] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communication sector

[12] Directive 2006/24/EC of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC

[13] *FAQ 2000-2007*, DataStorm Information Systems, Available at: http://whatismyipaddress.com/staticpages/index.php/FAQ

[14] Fiala, J. et al. 1993, *Občanské právo hmotné*, Masarykova univerzita, Brno.

[15] Latanicki, J. 2007, *The dark side of new technologies, Privacy and Trust (presentation)*, Thales European REsearch centre for Security & Information Systems

[16] Lloyd, I. J. 2000, *Information Technology Law*, Butterworths, London, Edinburgh, Dublin

[17] LoPucki, L. 2001, Human Identification Theory and the Identity Theft Problem, *Texas Law Review*, vol. 80: 89, pp. 89-135, Available at: http://ssrn.com/abstract=263213

[18] Polčák, R. et al. 2007, *Introduction to ICT Law (Selected Issues)*, Masarykova univerzita, Brno

[19] Post, D. G. 1996, Pooling Intellectual Capital: Thoughts on Anonymity, Pseudonymity, and Limited Liability in Cyberspace, *The University of Chicago LEGAL FORUM*, vol. 1996, pp 139-169.

[20] Preneel, B. 2006, *An Introduction to Modern Cryptology*

[21] Reed, Ch. 2000, *Internet Law: Text and Materials*, Butterworths, London, Edinburgh, Dublin

[22] Skinner, G., Han, S., Chang, E. 2006, An information privacy taxonomy for collaborative environments, *Information Management & Computer Security*, vol. 14, no. 4, pp. 382-394, Available at: http://www.emeraldinsight.com/Insight/ViewContent Servlet?Filename=Published/EmeraldFullTextArticle/Articles/0460140406.html

[23] Smith, A. G. 1997, Testing the Surf: Criteria for Evaluating Internet Information Resources, *The Public-Access Computer Systems Review*, vol. 8, no. 3, Available at: http://epress.lib.uh.edu/pr/v8/n3/smit8n3.html

[24] Van Kokswijk, J. 2003, *Hum@n: Telecoms and Internet as Interface to Interreality, a search for adaptive technology and defining users*, Bergboek, Eindhoven