enisa
European Network and Information Security Agency

The right to be forgotten – between expectations and practice

Contributors to this report

Authors:

- Peter Druschel (Max Planck Institute for Software Systems, Germany)
- Michael Backes (Saarland University, Germany)
- Rodica Tirtea (ENISA)

ENISA project management:

- Rodica Tirtea
- Demosthenes Ikonomou

Agreements or Acknowledgements

- The authors would like to thank Prof. Bart Preneel (K. U. Leuven, Belgium) and Hannes Tschofenig (NSN, Finland) for their valuable feedback and comments on this document.

[Deliverable – 2011-10-18]

About ENISA

The European Network and Information Security Agency (ENISA) is a centre of network and information security expertise for the EU, its member states, the private sector and Europe's citizens. ENISA works with these groups to develop advice and recommendations on good practice in information security. It assists EU member states in implementing relevant EU legislation and works to improve the resilience of Europe's critical information infrastructure and networks. ENISA seeks to enhance existing expertise in EU member states by supporting the development of cross-border communities committed to improving network and information security throughout the EU. More information about ENISA and its work can be found at www.enisa.europa.eu.

Contact details

For contacting ENISA or for general enquiries on this subject, please use the following details:

- Email: sta@enisa.europa.eu
- Internet: http://www.enisa.europa.eu

For questions related to this project, please use the following details:

- Email: sta@enisa.europa.eu

Contents

1 Executive Summary

The right to be forgotten is included in the proposed regulation on data protection published by the European Commission in January 2012[1]. The regulation is still to be adopted by the European Parliament for entering into force. The different legal aspects of the right to be forgotten (i.e. right to erasure or right to oblivion) have been debated in different contexts[2] and are beyond the scope of this paper. With this paper we aim to cover other facets of the right to be forgotten. We focus on the technical means to enforce or support the right in information systems; as can be seen from this paper, there are technical limitations and there is a further need for clear definitions and legal clarifications.

This paper complements two other recent publications of ENISA: the study on data storage and collection in Europe, which, based on a survey of all 27 Member States, clarifies the complex landscape of approaches and practices in the EU; and, the paper on online behavioural tracking, which shows the extent of online profiling and its privacy implications.

A unified and uniform approach is needed and desired in Europe to be able to secure the fundamental right for personal data protection. The reform of the data protection laws in Europe is a clear step in this direction and through its' work, ENISA provides a technical perspective, supporting the reform by bringing forward an information security perspective.

Information security technology plays critical role in enforcing the right to be forgotten. In this paper we review relevant existing technology, and identify technical limitations and challenges when trying to enforce the right. Furthermore, we identify the need for additional definitions and legal clarifications required before appropriate technical means to enforce the right can be implemented.

The recommendations of the paper cover multiple aspects:

- Technical means of assisting the enforcement of the right to be forgotten require a definition of the scope of personal data, a clarification of who has the right to ask for the deletion of personal data under what circumstances, and what are acceptable ways to affect the removal of data. Data Protection Authorities, the Article 29 Data Protection Working Party, the European Data Protection Supervisor, etc. should work together to clarify these issues. Furthermore, when providing the abovementioned definitions, the technical challenges in

---

[1] *European Commission, Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final, 25 January 2012, available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf (last accessed on 28.03.2012).*

[2] *Ambrose, Meg Leta and Ausloos, Jef, The Right to Be Forgotten Across the Pond (March 31, 2012). 2012 TRPC. Available at SSRN: http://ssrn.com/abstract=2032325 or http://dx.doi.org/10.2139/ssrn.2032325 , last visited September 2012.*

enforcing the right to be forgotten (and the associated costs) for a given choice of definition should be considered carefully.

- For any reasonable interpretation of the right to be forgotten, a purely technical and comprehensive solution to enforce the right in the open Internet is generally impossible. An interdisciplinary approach is needed and policy makers should be aware of this fact.
- A possible pragmatic approach to assist with the enforcement of the right to be forgotten is to require search engine operators and sharing services within the EU to filter references to forgotten information stored inside and outside the EU region.
- Particular care must be taken concerning the deletion of personal data stored on discarded and offline storage devices.
- Data controllers should be required to provide users with easy access to the personal data they store and ways to update, rectify, and delete data without undue delay and without cost to the user (to the extent that this does not conflict with other applicable laws).
- Research communities, industry, etc. should develop techniques and coordinate initiatives that aim at preventing the unwanted collection and dissemination of information (e.g., robot.txt, do not track, access control).

As mentioned above, this paper is complementing two other recent publications of ENISA in this area. In this broader context, given the findings of this paper, ENISA recommends that policy makers should ensure the use of technologies supporting the principle of minimal disclosure in order to minimize the amount of personal data collected and stored online. We also recommend the use of encryption for the storage and transfer of personal data. Particular attention should be focusing on tracking and profiling online, and enforcement solutions should be deployed to block misbehaving players and to force compliance with rules and regulations regarding personal data protection.
At the same time, Data Protection Authorities, the Article 29 Data Protection Working Party, the European Data Protection Supervisor, etc. should work together to clarify pending definition issues taking into account the practical implementation aspects while Member States should eliminate conflicting regulations.

2 Introduction

The 'Right to be forgotten' and topics related to the 'minimal disclosure' and 'minimum duration of the storage of personal data' are covered by the data protection policy framework.

Policy context. The right to be forgotten is described in Article 17 of the Regulation proposed in January 2012[3] by European Commission to replace the existing Data Protection Directive. In Article 5(e) is specified that '*Personal data must be […] kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed […]*[4]*'*. The Article 17, '*Right to be forgotten and to erasure*', provides the conditions of the right to be forgotten, including the obligation of the controller who has made the personal data public to inform third parties on the data subject's request to erase any links to, or copy or replication of that personal data[5]. Preamble (53) specifies: '*Any person should have the right to have personal data concerning them rectified and a 'right to be forgotten' where the retention of such data is not in compliance with this Regulation. In particular, data subjects should have the right that their personal data are erased and no longer processed, where the data are no longer necessary in relation to the purposes for which the data are collected or otherwise processed, where data subjects have withdrawn their consent for processing or where they object to the processing of personal data concerning them or where the processing of their personal data otherwise does not comply with this Regulation.'* And in (54) *'To ensure this information, the controller should take all reasonable steps, including technical measures, in relation to data for the publication of which the controller is responsible. In relation to a third party publication of personal data, the controller should be considered responsible for the publication, where the controller has authorised the publication by the third party'.*

Implementing acts are still needed to clarify how this important right will be implemented. In this paper we are identifying some technical challenges related to this objective.

---

[3] *European Commission, Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final, 25 January 2012, available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf (last accessed on 28.03.2012)*

[4] *Personal data may be stored for longer periods insofar as the data will be processed solely for historical, statistical or scientific research purposes in accordance with the rules and conditions of the proposed regulation and further delegated acts.*

[5] *In the same Article is mentioned, that if the controller has made the personal data public, it 'shall take all reasonable steps, including technical measures, in relation to data for the publication of which the controller is responsible, to inform third parties which are processing such data, that a data subject requests them to erase any links to, or copy or replication of that personal data. Where the controller has authorised a third party publication of personal data, the controller shall be considered responsible for that publication.'*

Related work at ENISA. This paper is complementing the '*Study on data collection and storage in the EU*[6]*'* conducted during 2011 and published in 2012 by ENISA and the paper on privacy considerations of online behavioural tracking[7].

The study on data storage and collection covers an analysis of the relevant legal framework of European Member States on the *principle of minimal disclosure* and the *minimum duration of the storage of personal data*. The study surveys the implementation of the minimal disclosure principle, using 3 case studies (social networks, public transportation and telecom services) across 27 Member States. The findings show that more work is needed for the principle to be deployed correctly and there is a need for clear enforcement mechanisms.

Furthermore, the above mentioned study identifies issues from a pan European perspective regarding the storage of personal data by telecom operators, one of the more regulated industries (see section 5.3.2 'Storage of personal data by telecommunication operators' of the study). The study shows that the Member States have adopted varying practices with regard to the implementation of this storage obligation, the majority of the Member States have literally transposed the provision of the ePrivacy Directive and allow the processing of traffic data necessary for the purposes of billing and interconnection payments only up to the *end of the period during which the bill may lawfully be challenged or payment pursued*, without specifying in their national law how long this period would be. In practice the storage period is between 3-6 months to 10 years across 27 Member States.

The study concludes with a set of recommendations; the most relevant for this paper is the one addressed to the Member States that they should identify and eliminate conflicting regulatory provisions relating to the collection and storage of persona data. Furthermore, the study recommends to support minimal data disclosure and to encourage minimal data storage periods: *the practical implementations of the principles of data minimisation and of conservation in specific cases by data controllers, should be evaluated (for instance in the form of audits) and clear sanctions and enforcement mechanisms should be available in cases of violations.*

Among the recommendations of the paper on online tracking[8], we select, as relevant also for this study, the following: more work and an interdisciplinary approach are needed to address the privacy risks associated with tracking mechanisms. There is a further need to development anti-tracking initiatives and solutions; development of easy-to-use tools for transparency and control; awareness is important but there is a need to enhance transparency tools to allow the users to know how their

---

[6] *ENISA, Study on data collection and storage in the EU, available at:* http://www.enisa.europa.eu/act/it/library/deliverables/data-collection

[7] *'Privacy considerations of online behavioural tracking', paper to be published by ENISA under* http://www.enisa.europa.eu/act/it/library/ .

[8] *Privacy considerations of online behavioural tracking', to be published by end of October 2012 by ENISA.*

personal data is collected, managed and transferred. Enforcement solutions should be deployed to block misbehaving players and to force compliance with rules and regulations regarding personal data protection; mechanisms should be defined by regulatory bodies both for compliance and for monitoring and detection of violation of the rules.

Structure of this paper. In this paper we focus on the technical means to achieve forgetfulness in information systems. After a discussion of possible definitions and interpretations of the right to be forgotten and their impact on possible technical implementations, the paper reviews some of the available techniques that allow expiration of electronically stored information. Challenges and vulnerabilities are summarized. The paper ends with a conclusions and recommendations section.

3 Interpreting the right to be forgotten

The right to be forgotten is subject to interpretation, as it does not provide precise definitions for what constitutes personal data, who has the right to request deletion of a particular data item, and what are acceptable means of implementing the right. It is up to the courts to interpret the law in ways appropriate to specific cases and evolve the case law as new applications, products and scenarios emerge.

However, different interpretations and definitions influence the technical and non-technical challenges in supporting the right to be forgotten in a fundamental way. Here, we briefly discuss the dimensions of possible definitions and interpretations, in order to provide background for the discussion of technical means in the following section.

3.1 What is the scope of personal data?

The new proposed EU regulations define personal data in art 4 as follows: *"(1) 'data subject' means an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person; (2) 'personal data' means any information relating to a data subject."*

Data protection directive[9], definitions in art. 2 are *"(a) 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); 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."*

These definitions define personal data broadly as information that can be linked, either by itself or in combination with other available information, to uniquely identify a natural person. However, they leave to interpretation whether it includes information that can be used to identify a person with high probability but not with certainty, e.g. a picture of a person or an account of a person's history, actions of performance. Neither is it clear whether it includes information that identifies a person not uniquely, but as a member of a more or less small set of individuals, such as a family.

A related question is how aggregated and derived forms of information (e.g. statistics) should be affected when some of the raw data from which statistics are derived are forgotten. Removing forgotten information from all aggregated or derived forms may present a significant technical challenge. On the other hand, not removing such information from aggregated forms is risky, because it may be possible to infer the forgotten raw information by correlating different aggregated forms.

---

[9] *DPD available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML*

The difficulty is that the EU regulations and laws tend to be deliberately broad and general, to allow for a range of interpretations appropriate for many different situations. However, *technical means to ensure the right to be forgotten require a precise definition of the data and circumstances to which the right to be for forgotten shall apply*.

3.2 Who has the right to request deletion of a data item?

Next, we consider the question of who has the right to request deletion of a data item. In many cases, the answer is unambiguous, such as when a person requests that their own name, date-of-birth and residential address are removed from a database. In other cases, however, the question of who has the right to demand that an item should be forgotten is subject to interpretation.

For instance, consider a photograph depicting Alice and Bob engaged in some activity at a given time and place. Suppose Alice wishes the photo to be forgotten, while Bob insists that it persist. Whose wishes should be respected? What if multiple people appear in a group photo? Who gets to decide if and when the photo should be forgotten?

In another example, Bob incorporates part of a tweet he receives from Alice into a longer blog post of his own. When Alice later exercises her right to remove her tweet, what effect does this have on the status of Bob's blog post? Does Bob have to remove his entire blog post? Does he have to remove Alice's tweet from it and rewrite his post accordingly? What criteria should be used to decide?

A related question is how the right to be forgotten should be balanced against the public interest in accountability, journalism, history, and scientific inquiry? Should a politician or government be able to request removal of some embarrassing reports? Should the author of a scientific study be able to request withdrawal of the publication? What principles should be used to decide, and who has the authority to make a decision?

3.3 What constitutes "forgetting" a data item?

Our next question concerns the question of what is an acceptable way of "forgetting" information. A strict interpretation would require that all copies of the data be erased and removed from any derived or aggregated representations to the point where recovering the data is impossible by any known technical means. A slightly weaker (and possibly more practical) interpretation would allow encrypted copies of the data to survive, as long as they cannot be deciphered by unauthorized parties. An even weaker (and more practical) interpretation would allow clear text copies of the data to survive, as long as the data would no longer appear in public indices, database query results, or in the results of search engines.

4 Technologies and challenges

The fundamental technical challenge in enforcing the right to be forgotten lies in

(i) allowing a person to identify and locate personal data items stored about them;

(ii) tracking all copies of an item and all copies of information derived from the data item;

(iii) determining whether a person has the right to request removal of a data item; and,

(iv) effecting the erasure or removal of all exact or derived copies of the item in the case where an authorized person exercises the right.

In a completely open system like the (vast) public portion of today's world-wide web, anyone can make copies of a public data item and store them at arbitrary locations. Moreover, the system does not account for the number, owner or location of such copies. *In such an open system it is not generally possible for a person to locate all personal data items (exact or derived) stored about them; it is difficult to determine whether a person has the right to request removal of a particular data item; nor does any single person or entity have the authority or jurisdiction to effect the deletion of all copies*. Therefore, enforcing the right to be forgotten is impossible in an open, global system, in general.

The ability to enforce a "right to be forgotten" crucially depends on the capabilities of the underlying information system. In a nutshell, this capability is technically feasible only in "closed" systems, which reliably account for the processing, storage and dissemination of all information, and prevent the dissemination of data to locations where an erasure cannot be enforced. In such a system, all participating entities must reside in a jurisdiction that enforces the right to be forgotten, every data request must be authenticated and logged, and the principals must be linkable to real-world persons or organizations.

In principle, systems such as corporate networks and access-controlled public networks that fall entirely within the jurisdiction of EU member states, could meet these requirements. However, such networks would require, without exception, that all principals (users and providers) be strongly authenticated using a form of electronic identity that can be linked to natural persons.

In an open system such as the public portion of the Internet, on the other hand, public data can be accessed by principals with online identities that cannot be reliably linked to a natural person. These principals are capable of further distributing the information to other untrusted parties, possibly resulting in a massive replication of data. In such a system, there is no generally applicable, technical approach to enforce the right to be forgotten. This case is common in the Internet, e.g., when personal data is being included in social networking sites, homepages, blogs, tweets, etc. In the following subsections, we discuss both scenarios in more detail.

It is important to understand *that regardless of the type of information system, unauthorized copying of information by human observers is ultimately impossible to prevent by technical means*. Consider

Alice viewing Bob's personal information on a computer screen, while she is allowed to do so (i.e., before Bob has invoked his right to be forgotten). Alice can take a picture of the screen using a camera, take notes or memorize the information. It is technically impossible to prevent Alice from doing so, or even to recognize that she has obtained a copy of Bob's personal data. Later, when Bob invokes his right to be forgotten, all known copies of his data within the system are deleted. However, Alice has a copy of the information and she can distribute or re-publish this information at will.

4.1 Closed systems

For the purposes of this discussion, a closed system is one in which all components that process, transmit or store personal information, as well as all users and operators with access to personal information can be trusted or held accountable for respecting applicable laws and regulations concerning the use of the private information.

An example is a corporate network, where personal data is processed, transmitted and stored exclusively by data processing hardware and software owned and operated by the corporation, and all users and operators with access to personal information are employees. Implementing the 'right to be forgotten' in such a network is technically feasible, though not without its challenges. For instance, when the owner of an item of personal information exercises her right to be forgotten, finding and removing all copies of the information and any derived information (including cached copies stored on the local disks of employee computers, backup copies stored on archival storage media, etc.) can be technically challenging and require substantial operational overhead.

A more complex type of closed system is an industry that shares personal information and is regulated by the government regarding the use of this information. For instance, the United States health care industry (health providers, insurance companies, and health care billing companies) as well as employers share patient records, and are jointly responsible for handling this information in accordance with Title II of the US Health Insurance Portability and Accountability Act (HIPAA). Participating companies and organizations are trusted and held accountable for their appropriate use of personal information. The system is closed because all parties with access to the personal information are held accountable for their compliance with the law, and all personal information remains with the jurisdiction of the USA. In principle, the "right to be forgotten" can be implemented in such a system. In practice, however, privacy breaches in the healthcare sector are not uncommon, nor are losses of credit card information in the banking industry. This suggests that enforcing the right to be forgotten may be challenging even in closed systems.

Even in closed systems, *all users with access to personal information must be trusted to respect applicable privacy laws*, because it is very difficult to ensure compliance by technical means alone. For instance, it would be difficult to prevent an employee from using his smartphone to take pictures of personal information on his office computer screen and transport the digitized information outside the company, where they would be out of reach of any technical enforcement of the right to be forgotten.

4.2 Open systems

In open networks such as the Internet, *information accessible to the public typically cannot be kept under the control of the user who originated the data. The reason is that data can be digitally copied, stored locally, and re-entered into the Internet, often in different locations for different purposes*.

Such digital copying and re-insertion of arbitrary data cannot be generally prevented by technical means, unless one is able to make very strong assumptions about the underlying software and/or hardware, as in Digital Rights Management (DRM). Such strong assumptions introduce additional technical and economic challenges, and often meet with limited public acceptance. Moreover, it remains unclear if even these strong measures can solve the problem entirely. For instance, digital rights management requires a cryptographic infrastructure to protect the desired content, and the corresponding software programs have to be tailored to support DRM. Nevertheless, expert attackers can circumvent DRM with modest effort.

A potential solution to avoid the unauthorized duplication of data would be to augment data with an executable program that enforces copyright protection. For instance, images could be equipped with a corresponding displaying program that, e.g., communicates with a trusted server to properly display the (formerly encrypted) data, disables screen shots of images while the images are being viewed, and so forth. *These techniques could be used to hamper the duplication of data. However, this solution faces important limitations in practice:*

- First, virtually all services rely on standard file formats such as JPG, and hence would not accept proprietary formats that come with their own interpreter.
- Second, such solutions would often require additional communication with external servers, which raises additional security challenges. For instance, such programs would provide a new avenue for the injection of Trojan horses and viruses on individual's computers and devices. In order to function properly, these programs would have to execute with generous permissions, which could be exploited by potentially malicious code.

As a consequence, such solutions would likely meet with scepticism by industry, security experts and the public.

It is thus fair to say that *digital duplication cannot be prevented in general in open networks*. However, it is worth pointing out that even if one assumed that direct digital duplication can be excluded by technical means, there exist additional ways to effect data duplication which are even harder to prevent. For instance, taking a photograph of the screen while personal data is being displayed, or re-recording a private conversation using a microphone while it is being replayed cannot be prevented without physical screening of all listeners. Such re-captured information can be reinserted in the Internet.

Finally, truly public information such as important news typically exist in a variety of different forms, both in various digital places, as well as in non-digital newspapers, radio, etc. There is no technical way to make this data forgotten.

4.3 Protecting personal data on discarded or offline storage equipment

Regardless of whether private data is processed in an open or closed system, particular attention must be paid to personal data stored on discarded storage equipment, e.g., the magnetic and flash disk devices in discarded or recycled smart phones, notebooks, desktop computers and USB sticks. Simply deleting the files on such devices is insufficient to prevent third parties from recovering such data from discarded devices using simple and widely available technical means.

Based on the study presented by the ICO (Information Commissioner's Office – UK) in April 2012 [13], leaking personal data on discarded storage devices is a significant issue in practice, which consumers must take more care to avoid. The problem can be avoided by physically destroying the storage media, disposal by a trustworthy professional service, or the use of secure scrubbing software like DBAN [14].

A related question concerns that removal of data stored on off-line storage media. Storing data copies on off-line media like tape archives or USB sticks is common and required for disaster recovery. Locating such copies as part of a removal operation when a person exercises their right to be forgotten, however, can be particularly challenging. It may be necessary to distribute and retain removal requests indefinitely, so that removed data items stored on off-line media can be deleted as soon as the media is connected.

4.4 Existing techniques for expiration of data

Meyer-Schönberger [1] discusses the broad consequences of the virtually unbounded memory in current information systems. He proposes to tag sensitive data with an expiration date and to require all servers handling such data to obey that date. Such a mandate conflicts with the business interests of many corporate information service providers, and with the interests of governments. A word-wide legal mandate to respect data expiration dates, on the other hand, seems out of reach in the foreseeable future. Moreover, it is technically difficult to verify that all servers actually delete the data and hold delinquent service providers accountable. In recent years, a number of research projects have sought to address various aspects of this problem.

In a first line of work, the expiration date associated with personal data is implemented by encrypting the data itself with a symmetric encryption key and restricting access to these keys. Unfortunately, none of these systems are capable of dealing with scenarios where published data is subsequently manipulated, e.g., re-encoding of JPEGs as commonly done in social networks such as Facebook. Similarly, the challenges of storing and distributing keys securely during their period of validity has not

been considered thus far. Finally, preventing unauthorized copying, notifying all key holders to delete their keys, and authenticating removal requests remain open problems.

One line of work is aimed at effectively closed systems, such as corporate information systems and systems limited to a single jurisdiction. Here, all servers are aware of the encrypted nature of the data, post-processing of data is not supported, and the threat of an adversary who obtains keys while the data is available is generally not considered. The first such system we are aware of is described in [2], which provided the basic principle. Another prominent system is the *Ephemerizer* [3], an improved version of which was later described in [4].

Vanish [5] is a more recent approach, which stores shares of the keys in a distributed hash table (DHT), a data structure that underlies P2P networks. The DHT is designed to reliably remove the key after a certain time, so that the cleartext data becomes unavailable (although encrypted copies may persist). An attack against the proposed implementation was recently published [6], using a Sybil attack on the DHT.

X-pire! [7] is a system that allows users to set an expiration date for images in social networks (e.g., Facebook and Flickr) and on static websites, without requiring any form of additional interaction with these web pages. Technically, X-pire! encrypts images in a suitable manner before they are uploaded to the web server, and stores the corresponding keys on a dedicated key server. If a user Bob wishes to view this image, e.g., while visiting Alice's Facebook profile, a browser plugin executing on Bob's machine requests the corresponding decryption key from the key server. If the key has not yet expired, the image is decrypted and displayed on Bob's screen. As with all such systems, an attacker can obtain an unauthorized copy of the image by taking a screen snapshot or taking a picture with a camera while the cleartext image is being displayed on the screen.

The *EphCOM* system [8] is similar to Vanish, but uses a trick to store the keys in the cache of DNS servers, based on the presence of generated hostnames. Similar to Vanish and X-pire!, post-processing of protecting data and the threat of retrieving keys during their period of validity is not considered. To publish data that expires at a specific time one needs to find a (large) number of domains that have the same TTL; a study shows that TTLs of more than 7 days are rather uncommon, which limits the practicality of the approach.

A second line of work aims at securing privacy-sensitive content published in social networks, and is based on a different assumption: The central difference is that these approaches store all data on an external, trusted server, while the aforementioned approaches host these data on the target servers / the actual social networks themselves. One example is *FaceCloak* [9]; similar approaches are described in [10] and [11]. A challenge for these approaches is the scalability of a centralized server, given the vast amount of data, including multimedia data like images and videos published every day.

Following a similar motivation of constraining access to personal information, an owner-centric architecture (OCN) [12] has recently been proposed, which considers content ownership as the central principle. As far as content distribution and control is concerned, OCN aims to establish dedicated

storage locations for all data, to which only the legitimate owner of the data has access. Content access to such data is given to users by providing links to these storage places. The architecture ensures control of the data by its legitimate owner as long as users do not create unauthorized copies. Similar to the aforementioned approaches, duplication and uncontrolled dissemination of data would invalidate the guarantees offered by this architecture. As discussed in the preceding sections, these solutions thus only offer a partial technological solution to realize the right-to-be-forgotten.

Another potential solution to preventing data duplication is to adopt techniques from *Digital Rights Management (DRM)*. Technically, DRM is a technology for access control, which has been used by various commercial entities to inhibit, unsuccessfully, unauthorized usage of digital content. DRM was designed for content publishers, hardware manufacturers, copyright holders, and individuals who wish to limit the dissemination of their product after sale. In practice, DRM has faced various problems, especially at the technical level. DRM techniques can often be bypassed with moderate effort. For instance, DRM audio protection can be bypassed by burning audio files on an audio CD, and subsequently ripping them into DRM-free audio files. Similarly, watermarks in images and video can often be easily removed. As a consequence, solutions based on DRM are unlikely to be able to avoid data duplication in general.

While it is impossible in general to remove data from the Internet once it was published, it might be possible to limit its accessibility. One such approach relies on the observation that users typically find information on the Internet by issuing queries to a search engine, or by using a social networking, sharing, or tagging site. Data not identified by a search engine or shared via a service like Twitter is difficult to find. A natural way to "mostly forget" data is thus to prevent its appearance in the results of search engines, and to filter it from sharing services like Twitter. EU member states could require search engine operators and sharing services to filter references to forgotten data. As a result, forgotten data would be very difficult to find, even though copies may survive, for instance, outside the EU jurisdiction.

To summarize, *all existing technical approaches to ensure the right to be forgotten are vulnerable to unauthorized copying while the date is publicly accessible and a re-dissemination of such unauthorized copies once the data has expired*. Therefore, *the right to be forgotten cannot be ensured using technical means alone*. A possible partial solution may be a legal mandate aimed at making it difficult to find expired personal data, for instance, by requiring search engines to exclude expired personal data from their search results.

5 Conclusions and recommendations

Once personal information is published, it is ultimately impossible to prevent, or even observe, by technical means, the creation of unauthorized copies of this information. In an open system like the Internet, the right to be forgotten cannot be enforced by technical means alone. Enforcement must rest on a combination of technical and international legal provisions.

Recommendations:

- Technical means of assisting the enforcement of the right to be forgotten require a definition of the scope of personal data, a clarification of who has the right to ask for the deletion of personal data under which circumstances, and what are acceptable ways to affect the removal of data. Data Protection Authorities, the Article 29 Data Protection Working Party, the European Data Protection Supervisor, etc. should work together to clarify these issues.
- When providing the abovementioned definitions, the technical challenges in enforcing the right to be forgotten (and the associated costs) for a given choice of definition should be considered carefully.
- For any reasonable interpretation of the right to be forgotten, a purely technical and comprehensive solution to enforce the right in the open Internet is generally impossible.
- A possible pragmatic approach to assist with the enforcement of the right to be forgotten is to require search engine operators and sharing services within the EU to filter references to forgotten information stored inside and outside the EU region.
- Particular care must be taken concerning the deletion of personal data stored on discarded and offline storage devices.
- Data controllers should be required to provide users with easy access to the personal data they store and ways to update, rectify, and delete data without undue delay and without cost to the user (to the extent that this does not conflict with other applicable laws).
- Develop techniques that aim at preventing the unwanted collection and dissemination of information (e.g., robot.txt, do not track, access control).

As mentioned already, this paper is complementing two other recent publications of ENISA in this area. In this broader context, ENISA recommends

- To policy makers should ensure the use of technologies supporting the principle of minimal disclosure in order to minimize the amount of personal data collected and stored online.

- We also recommend for all parties the use of encryption for the storage and transfer of personal data.
- Particular attention should be focusing on tracking and profiling online, and policy makers should provide clear sanctions and means for enforcement in order to block misbehaving players and to force compliance with rules and regulations regarding personal data protection.
- The Data Protection Authorities and relevant stakeholders in the field should aim to improve user awareness relating to their rights stemming from the data protection legislation and on the possibilities offered to them by the legal system to exercise these rights, including by complaining in cases of excessive collection and storage of personal data.
- At the same time, Data Protection Authorities, the Article 29 Data Protection Working Party, the European Data Protection Supervisor, etc. should work together to clarify pending definition issues taking into account the practical implementation aspects while Member States should eliminate conflicting regulations (the collection and storage of personal data is not always only governed by the data protection legislation).

6 Bibliography

[1] V. Mayer-Schoenberger. Useful Void: The Art of Forgetting in the Age of Ubiquitous Computing, KSG Working Paper No. RWP07-022. Available at SSRN: http://ssrn.com/abstract=976541, year = 2007.

[2] D. Boneh and R. J. Lipton. A revocable backup system. Proc. Usenix Security, 91-96, 1996.

[3] Radia Perlman. File System Design with Assured Delete. Proc. 3rd IEEE International Security in Storage Workshop (SISW 2005), 2005.

[4] S. Nair, M. Dashti, B. Crispo, and A. Tanenbaum. A Hybrid PKI-IBC Based Ephemerizer System. Proc. New Approaches for Security, Privacy and Trust in Complex Environments, 241-252, 2007.

[5] R. Geambasu, A. Levy, T. Kohno, A. Krishnamurthy and H. M. Levy. Comet: An Active Distributed Key-Value Store. Proc. OSDI, 2010.

[6] S. Wolchok, O. S. Hofmann, N. Heninger, E. Felten, J. A. Halderman, Chr. Rossbach, B. Waters, and E. Witchel. Defeating Vanish with Low-Cost Sybil Attacks Against Large DHTs. Proc. NDSS, 2010.

[7] J. Backes, M. Backes, M. Duermuth, S. Gerling, and S. Lorenz. X-pire! - A digital expiration date for images in social networks. Available online at http://arxiv.org/abs/1112.2649, 2012.

[8] The Ephemeral Data Project. EphCOM: Practical Ephemeral Communications (How to implement ephemeral data with only primary Internet services). Available online at http://arxiv.org/abs/1003.5510, 2012.

[9] W. Luo, Q. Xie, and U. Hengartner. FaceCloak: An Architecture for User Privacy on Social Networking Sites. Proc. of 2009 IEEE International Conference on Privacy, Security, Risk and Trust (PASSAT 2009), 26 - 33, year = 2009.

[10] M. Lucas and N. Borisov. FlyByNight: mitigating the privacy risks of social networking. Proc. ACM workshop on Privacy in the electronic society, 1 - 8, 2008.

[11] S. Guha, K. Tang, and P. Francis. NOYB: Privacy in Online Social Networks. Proc. ACM SIGCOMM Workshop on Online Social Networks (WOSN), 2008.

[12] C. Castelluccia, E. De Cristofaro, A. Francillon, M. A. Kaafar "EphPub: Toward Robust Ephemeral Publishing" In Proceedings of the 19th IEEE International Conference on Network Protocols (ICNP 2011), 2011.

[13] UK Information commissioner's office. On: Deleting your data from computers, laptops and other devices. http://www.ico.gov.uk/for_the_public/topic_specific_guides/online/deleting_your_data.aspx

[14] Darik's boot and nuke (DBAN). http://www.dban.org/

P.O. Box 1309, 71001 Heraklion, Greece
www.enisa.europa.eu