Protection of personal data in security alert sharing platforms

Václav Stupka
Institute of Computer Science
Masaryk University
Brno, Czech Republic
stupka@ics.muni.cz

Martin Horák
Institute of Computer Science
Masaryk University
Brno, Czech Republic
horak@ics.muni.cz

Martin Husák
Institute of Computer Science
Masaryk University
Brno, Czech Republic
husakm@ics.muni.cz

ABSTRACT

In order to ensure confidentiality, integrity and availability (so called CIA triad) of data within network infrastructure, it is necessary to be able to detect and handle cyber security incidents. For this purpose, it is vital for Computer Security Incident Response Teams (CSIRT) to have enough data on relevant security events and threats. That is why CSIRTs share security alerts and incidents data using various sharing platforms. Even though they do so primarily to protect data and privacy of users, their use also lead to additional processing of personal data, which may cause new privacy risks. European data protection law, especially with the adoption of the new General data protection regulation, sets out very strict rules on processing of personal data which on one hand leads to greater protection of individual's rights, but on the other creates great obstacles for those who need to share any personal data. This paper analyses the General Data Protection Regulation (GDPR), relevant case-law and analyses by the Article 29 Working Party to propose optimal methods and level of personal data processing necessary for effective use of security alert sharing platforms, which would be legally compliant and lead to appropriate balance between risks.

CCS CONCEPTS

• **Security and privacy** → *Intrusion/anomaly detection and malware mitigation*; • **Social and professional topics** → *Privacy policies*;

KEYWORDS

Intrusion detection, Information sharing, Alert sharing platform, Personal data, Privacy, Cyber security

ACM Reference format:
Václav Stupka, Martin Horák, and Martin Husák. 2017. Protection of personal data in security alert sharing platforms. In *Proceedings of ARES '17, Reggio Calabria, Italy, August 29-September 01, 2017,* 8 pages.
DOI: 10.1145/3098954.3105822

1 INTRODUCTION

Frequency and magnitude of cyber security incidents is constantly increasing. Due to the borderless nature of these incidents, it is absolutely necessary for Computer Security Incident Response Teams (CSIRT) to coordinate their efforts, ensure rapid response and share data regardless of what kind of organisation they aim to protect. For this purpose they use wide spectrum of technical tools from which are some of the most effective alert sharing platforms. They allow for effective sharing of information about cyber security incidents and threats and their comprehensive analysis. Use of these platforms not only protects peers of the platform, but also helps to increase security of the whole cyberspace.

Some of the data shared and analysed using these platforms however contain information about users or other persons. This personal data is in most modern countries protected by data protection laws, because in some cases can its use and processing threaten privacy of data subjects. European union is at the forefront of legal protection of privacy, since it strictly regulates processing of personal data. The current EU 1998 Data Protection Directive will be replaced in May of 2018 by General Data Protection Regulation (GDPR) which imposes even more strict rules to handling personal data. In order to ensure legality of alert sharing platforms, it is necessary to make them compliant with the GDPR. In this paper, we address the issue of compliance. The research question is:

> **How the data protection limit the operation of security alert sharing platforms and how to ensure its compliance?**

This paper is organized into six sections. Section 2 surveys related work. Section 3 provides technical background on security alert sharing platforms, while legal background is provided in Section 4. Proposed measures on how to make security alert sharing platforms compliant with GDPR are presented in Section 5. Section 6 concludes the paper.

2 RELATED WORK

The related work consists of two parts, technical background and legal aspects.

2.1 Technical background

The alert sharing platforms emerged from two theoretical concepts, alert correlation and collaborative intrusion detection. The approaches to alert correlation were proposed by Valeur et al. [19]. In this work, alert life-cycle and common tasks of alert correlation are proposed. Collaborative security and collaborative intrusion detection were surveyed recently by Meng et al. [15] and Vasilomanolakis et al. [21]. Formats and protocols for exchanging security events were briefly surveyed by Steinberger et al. [18].

As the research continues, the CSIRT community makes efforts to implement collaborative security and information sharing in practice. These efforts were surveyed in reports by ENISA [9]. Unfortunately, many competing standards, tools, and platforms exist and it is hard to select the outstanding ones. One of the most well-known information sharing platforms is MISP [22], that focuses on exchange of indicators of compromise (IoC) obtained from malware.

ARES '17, Reggio Calabria, Italy

An example of centralized alert sharing platform is CIF (Collective Intelligence Framework). Lately, TAXII (Trusted Automated eXchange of Indicator Information) gain a lot of attention. The comprehensive list of platforms is also provided by ENISA [10].

The data exchange formats and serialization are also worth mentioning as their many examples get a good insight into what is a security alert. IDMEF [7] is one of the most well-known security alert exchange formats that is used among the research community. Practitioners prefer IODEF [6], which is compatible with IDMEF. Lately, STIX gained attention in the cyber security community. There are a lot of examples of security alerts in these formats, as well as their comparison, in the related literature [10, 18].

2.2 Legal aspects

Legal aspects of alert sharing platforms belong to broader discussion about cyber security, privacy and law. Some of the basic concepts of this discussion can be found in the paper of Sokol et al. [17] In this paper there can be also found argumentation about IP addresses which authors regard as a personal data even in the meaning of the current Data Protection Directive.

Serrano et al. [16] focus on the legal aspects of the cyber security data sharing between organisations subject to different legal frameworks. They also propose a solution in form of Information Exchange Policy which is set up by the organisations themselves.

The European Network and Information Security Agency (ENISA) published study which deals with legal and regulatory factors and perform an assessment of what effects these factors have on cross-border data sharing between CERTs [8]. One these factors which the study focuses on is the European legal framework governing data protection and privacy. The study provides an example of legal checklist for privacy and data protection.

3 HOW THE SHARING PLATFORM WORKS

This section presents a background on security information sharing and alert sharing platforms. First, the purpose of sharing is stated, followed by the description of actual content, e. g., the shared data. Then, the life-cycle of data is described and the section closes with the benefits of shared data to their recipients.

3.1 What is the purpose of sharing

The motivation behind sharing the security alerts is an increased security among the peers. However, there are other purposes of sharing the data, e. g., building global cyber security situational awareness and cyber threat intelligence.

Isolated intrusion detection systems have only a limited scope of observation, e. g., on the network. They also cannot detect unknown attacks and they depend on updates from their vendors. Exchange of knowledge and information allows the peers to achieve higher accuracy in intrusion detection by correlating the results with others, receiving additional information and knowledge, and even preventing attacks that could otherwise remain unnoticed. Thus, collaboration increases the detection capabilities of individual peers [12].

The other use cases of sharing the cyber security data is building of global situational awareness and cyber threat intelligence. Keeping records on cyber events from a single observation point may be biased towards the systems and setup of a corresponding network. On the other hand, correlation of observations from multiple heterogeneous observation points allows better understanding of current trends in malicious network activities. Understanding which security events are of local significance and which are global allows us to distinguish targeted attacks.

3.2 What is being shared

Basically, every piece of data can be shared. However, the five most common types of shared data are reports of security incidents or alerts, indicators of compromise (IoC), raw data, and tools. Raw data are typically exchanged between collaborative intrusion detection systems [12]. "Incident" data usually contain all the information related to a security incident, including sensitive information, which cannot be shared easily. Security "events" contain non-sensitive metadata related to an incident [9]. Indicators of compromise are artifacts found in a system or in the network traffic that indicate an intrusion. The IoCs and security events are often interchangeable. Finally, the tools may also be a subject of sharing, but these typically do not contain any information apart from source codes or binaries.

The content of the shared data may vary depending on type of sharing and willingness of sharing peers to provide complete information. Raw data and incident data may contain basically everything, from malicious network traffic captures in a PCAP file to malware binaries and snapshots of infected systems. However, due to technical difficulties of sharing large volumes of data, only network traffic records, e. g., NetFlow [14], or system logs are being used. On the other hand, security events and IoC typically contain a small piece of information, typically IP addresses, email addresses, malware signatures, domain names, URLs, etc. The difference between security event and IoC contents are rather semantic, even if they are of the same type, e. g., IP address. Security events contain identifiers of active participants of an event, e. g., an attacker and a target. IoCs, on the other hand, contain information such as IP address of a botnet C&C center that is not yet necessarily involved in any activity.

3.3 Life-cycle of the data

There are many approaches to sharing the data. Thus, it is difficult to present a definite life-cycle of the data. For the purposes of this paper, we took the alert sharing platform SABU[1] as an example on which we show the alert life-cycle. Our alert sharing platform is centralized and distinguishes senders and receivers of the alerts. The data processing is performed at the central hub for all the alerts.

First, the data are collected. The alerts are generated by various intrusion and anomaly detection systems. Alternatively, the alerts can be inserted manually by users and obtained from another alert sharing platform.

Second, the data are distributed. In case of a centralized sharing platform, all the alerts are sent to a central repository. The central point then can process the data before their actual utilization. Alternatively, the data can be exchanged directly among peers, which is common in collaborative intrusion detection networks [12].

Third, the data are correlated. Alert correlation consists of many procedures that help in analyzing the data, but they also serve

[1] https://sabu.cesnet.cz/en/start

as means of data quality assurance. The common tasks of alert correlation, as described by Valeur et al. [19], are normalization, pre-processing, alert fusion, alert verification, thread reconstruction, focus recognition, impact analysis, and prioritization. The data are at first converted to a unified format, their syntax is validated, and the duplicated alerts are fused. Then, the alerts are mutually checked to confirm their sanity and to reveal false positives. Further, attack sequences are matched against historical records, and the anticipated target and impact are projected and the alerts are prioritized accordingly. In this phase of an alert life-cycle, the novel pieces of information are distilled from the alerts, often with a certain degree of uncertainty.

Fourth, the data are re-distributed and utilized by the receiving peers. In this phase, all the peers can access the data and utilize them, unless any restriction and specific distribution rules are used. Often the recipients use filters to receive only the alert of a certain type to avoid processing the large volumes of data. Once a recipients receive the data, they can be applied in protecting the network. Typically, the identifiers from the received alerts, e. g., IP addresses, are put on blacklists. Firewall rules and other network traffic filtering mechanisms may be applied to drop the incoming traffic from subjects enlisted in the alerts or the prevent local users from accessing them. In this phase, the data from the alerts are actively used to influence the outer world.

If a central repository is used, the data are kept there. It is vital to keep the data for longer time for the purposes of advanced alert analyses and threat intelligence. Permanent data storage can be used, but it is actually not common due to large volumes of the data (security alerts can be considered as big data) and short life-time of the individual alerts. Thus, the alerts are typically dismissed after several days or weeks. On the other hand, the derived and statistical information, e. g., counts and shares of alert types, are typically stored for a longer time.

3.4 How does the sharing benefit the data subject

Information sharing benefits the recipients of the data in multiple ways. The shared data may be used as a blacklist, early warning, and to improve accuracy and precision of intrusion detection.

The blacklists are reactive. They contain identifiers, e. g., IP addresses and URLs, of malicious entities that can be used for immediate incident response, e. g., network traffic filtering and access restrictions. This data should be highly trustful as they are used to protect the network of a recipient.

Early warning data are preventive and consist of events that are likely to happen or identifiers of subjects that are likely to be involved in future security events. For example, worm propagation was detection in one network and it is known from historical records that other networks are going to be infected soon, thus, the other networks are informed.

Finally, the shared data can be used by intrusion detection systems to increase capabilities and precision of intrusion detection. Shared IoCs and other identifiers and signatures allows for the detection of events that would otherwise remain unnoticed due to insufficient IDS capabilities.

4 LEGAL FRAMEWORK

4.1 Evolution of data protection in the EU

"Data protection" law refers to the legal scheme governing the collection, storage, processing, disclosure, and transfer of individual's personal data. Data protection is dynamic and increasingly important topic that is related to many core political and legal concerns, including the freedom of expression, security, and international business. Graham stated in one of his recent studies that "data privacy laws are spreading globally, and their number and geographical diversity accelerating since 2000." [13]

In Europe, which is recognized as a global leader in data protection, implemented legal tools focus on protection of individual's fundamental right to privacy in general, and more specifically on protection of right to information self-determination. These rights are in the EU guaranteed in two supranational conventions - the European Convention for Human Rights and the Charter of Fundamental Rights. In compliance with these conventions, individual's rights are enforced on three levels - by individual European states, by the Council of Europe and by the European Union. Anyone who seeks to collect, analyze or otherwise process the personal data of natural persons in Europe is required to take time to understand relevant European data protection rules.

Existing 1998 Data protection directive (no. 95/46/EC) sets out rules that are mandatory for EU member states to implement. This legal tool however seems to be rather obsolete for two main reasons. First, when the directive was drafted in 1995 legislators did not anticipate such a rapid development of information technologies. Second, it is implemented and interpreted in different EU states differently, which weakens legal certainty of data subjects and constitutes obstacles to the exercise of their rights. Until now, these shortcomings have been remedied by the case law, which has expanded the understanding of EU jurisdiction (e. g., in cases Google Spain C-131/12 and Weltimmo C-230/14) and the range of data that is considered personal (e. g., case Breyer C-582/14).

However, this was not enough, which is why the EU was from 2013 drafting a new data protection regulation that seeks to increase the level of data protection across the Union's twenty-seven member states and beyond. This regulation comes into force on 25 May 2018 as Regulation no. 2016/679. Another relevant piece of legislation is so called e-Privacy directive (no. 2002/58/EC) which focuses on the protection of privacy in the sector of electronic communications. This directive is a lex specialis to the Data protection directive (and to the GDPR), which deals specifically with privacy issues related to electronic communications, such as data retention, confidentiality of communications, etc. This directive will however also be subject to replacement by new regulation, proposal for which was introduced by European Commission in January 2017.

4.2 The GDPR

The GDPR replaces the current Data Protection Directive from 1998. The main intent of the GDPR is to give individuals more control over their personal data, impose stricter rules to companies handling it and make sure companies embrace new technology to process the influx of data produced.

The fact that it is a "regulation" instead of a "directive" means it is directly applicable to all EU member states without a need

for national implementing legislation, which will unify the rules through EU instead of merely harmonising them.

Basic principles of the data protection remain with the GDPR the same as before. It requires data controllers and processors to process personal data fairly, lawfully and transparently, to take all reasonable steps to ensure that personal data are accurate and secure. It also forbids them from using the data for different purposes, than for which it was collected, from storing data that are not needed to achieve its processing purposes and from storing the data longer than necessary in relation to the processing purposes.

Main changes that the GDPR introduces are following: First, it extends the jurisdiction of EU data protection law, as it applies to all entities processing the personal data of data subjects residing in the Union, regardless of the entities location. Second, it increases applicable sanctions, as organisations in breach of the GDPR can be fined up to 4% of annual global turnover or 20 million EUR (whichever is greater). Third, it introduces new rights of the data subject and related obligations for data controllers and processors.

Controllers will be obliged to notify the data subject of any data breach which is likely to "result in a risk for the rights and freedoms of individuals", to provide upon request from the data subject information about extent and purposes of processing of their data and dump of their personal data being processed, to delete upon request the personal data if it is no longer relevant to original purposes for processing, or if the data subject withdrawn the consent for processing.

The GDPR has also placed great emphasis on the accountability principle for data controllers and processors to demonstrate data compliance. They are newly required to maintain certain documentation, conduct impact assessment reports for riskier processing and employ data protection practices by default.

4.3 Personal data and it's use

The GDPR in its Art. 4 para. 1 defines the personal data very broadly as "any information about identified or identifiable person". This person is in the data protection terminology related to as the "data subject". The same article of the GDPR also states that by identifiable person is meant "a person, who may be directly or indirectly identified" and follows with demonstrative list of elements that may lead to identification including ID number, location data, online identifier or personal specifics.

For the purposes of alert sharing platforms is most relevant definition of online identifier. This term is further explained in recital 30 of the regulation, which states that "natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them." This concept in combination with relevant CJEU case-law, which applies rather extensive approach to the interpretation of the concept of personal data, leads to a situation where almost any data that is shared between peers of alert sharing platforms, could be personal data.

As we mentioned above in Section 3.2, any data that can peers get from the data traffic in their systems may be shared. Such data may contain not only identifiers like IP addresses, domain names, URLs, or email addresses, but also pieces of transferred content. Even considering the fact, that in some cases such data may not be sufficient to identify specific natural person and that individual peers will most likely share only limited amount of data to protect their users and themselves, we have to assume that this data, especially when combined, can directly or indirectly identify the data subject.

This approach is supported not only by Article 29 working party of data protection authorities of EU countries, which states in its opinion that "the controller that processes IP addresses anticipates that the "means likely reasonably to be used" to identify the persons will be available, e. g., through the courts appealed to (otherwise the collection of the information makes no sense), and therefore the information should be considered as personal data" [1], and also by CJEU case-law. In Scarlet case (No. C-70-10), the CJEU stated that "[IP] addresses are protected personal data because they allow users to be precisely identified" [4]. In Breyer case (no. c-582/14) the CJEU concluded that "dynamic IP address [...] constitutes personal data [...], in relation to that provider, where the latter has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person" [5], while by legal means might according to Sokol et. al. be for example just a possibility to hand the data over to the police which has then access to data retention data [17].

So it is safe to say that at least some of the data shared using alert sharing platforms should be considered personal data. At the same time, we have to take into account that we are not able to separate such personal data from other shared data, because it is technically impossible. Which is why all of the shared data should be treated as personal data, because, as Article 29 working party puts it, "unless the provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side" [1].

Now we know that we are dealing with personal data, but when and for what purposes can we use it? We mentioned purpose limitation above in the Section 4.2, this principle requires data controllers to collect personal data for specified, explicit and legitimate purposes. Such data may then be processed only in order to achieve declared purpose and if it is to be processed in a manner that is incompatible with this purpose, the data controller has to find another legal ground or cease the data processing.

The GDPR provides six legal grounds for personal data processing: consent of the data subject, performance of contract with data subject, legal obligation of the controller, protection of vital interests of the data subject, public interest and legitimate interest of the controller or third party. It would be practically impossible to get valid consent from every data subject, whose personal data are shared via alert sharing platforms, because in most cases the data controller is not in direct contact with the data subject. In specific cases may be the personal data shared using these platforms in compliance with legal obligation of the controller. For instance, according to so-called NIS directive (No. (EU) 2016/1148) requires

operators of essential services and digital services providers to notify competent authorities of any significant security incident, if they would do so using sharing platform, it would be legitimate and legal. However, in most cases are the alert sharing platforms operated by private entities and individual peers are not required by law to provide any data.

So we are down to the last two legal grounds - protection of vital interest of the data subject and legitimate interest of the controller or third party. Even though sharing of alert data aims to protect not only the sharing organisations, but also the security and privacy of users, we cannot assume that sharing of specific data is always done in order to protect its data subject. For instance if the data subject is attacker trying to hack into the computer system – here we are protecting users, but not the data subject.

4.4 Legitimate interest and proportionality test

The last legal ground is most fitting, but at the same time most complicated to use. According to the GDPR the processing of personal data is lawful if "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data" [3].

The GDPR provide for "legitimate interest" as a legal ground for processing personal data. Some cyber security professionals understood it so that their activities are always covered and they can freely use almost any personal data for security purposes without consent from data subjects. Even some EU member states saw it only as one of six options, and one which is no more or no less important than the other options, and which may apply in a large number and large variety of situations, provided the necessary conditions are met. This is however not always the case, because this legal ground should only be used sparingly to fill in gaps for rare and unforeseen situation as 'a last resort' - or as a last chance if no other grounds may apply. The Article 29 Working Party has already made it clear that merely having a legitimate interest by itself is not enough to entitle controller to use personal data. The objective of the "legitimate interest" provision is to provide them with "necessary flexibility for data controllers for situations where there is no undue impact on data subjects". The Article 29 Working Party also cautioned that this legal ground is not to be used "on the basis that it is less constraining than the other grounds" [1].

In comparison with the Data protection directive, the GDPR adds a requirement, that the 'legitimate interest' of the controller or third party can not justify use of personal data if their "interest is overridden by interest or fundamental rights of the data subject". Which means that a controller that plans to process personal data must balance its legitimate interest against the rights of the data subject, but also "the data subject's interests, irrespective of whether these interests are legitimate or not. Any controller that hopes to use legitimate interest also bears the onus for demonstrating that its interest is favored in such a balancing test" [2]. The Article 29 Working Party cautions that the balancing test should also be sufficiently documented in such a way that data subjects, data authorities, and the courts can examine.

In this balancing test, the controller that share large amount of data should, among other, consider, that if this data is combined with data shared by other peers then using big data analysis techniques might lead to unexpected results, which may be highly intrusive to the individual privacy.

Another factor in the balancing test is mentioned in Recital 47 of the GDPR: "...taking into consideration the reasonable expectation of data subjects based on their relationship to the controller". A controller involved in sharing of incident data must ask the following question: Is it reasonable to assume that a regular person who uses my infrastructure expects that their behaviour is being tracked and measured, consolidated across devices, and that the results of these operations are being traded between different organisations that he or she has never heard of, and retained for further trading and consolidation over considerable periods of time? [2]

Let's be more specific. If we want to justify processing and sharing of personal data in alert sharing platforms by "legitimate interest" of peers, we need to take following steps:

(1) determine, if this interest is legitimate one,
(2) determine whether the processing is necessary to achieve the interested pursued,
(3) establish a provisional balance by assessing whether the data controller's interest is overridden by the fundamental rights or interests of the data subjects,
(4) establish a final balance by taking into account additional safeguards,
(5) demonstrate compliance and ensure transparency.

In both the opinion of the Article 29 working party and the GDPR it is mentioned that cyber security is one of the fields, in which it is likely that the personal data will be processed due to legitimate interest of the controller. Recital 49 of the GDPR specifically states that "the processing of personal data to the extent strictly necessary and proportionate for the purposes of ensuring network and information security, [...] constitutes a legitimate interest of the data controller concerned" [3]. So it is probably safe to say that so called CIA triad[2] of data and information systems is a legitimate interest within the meaning of the GPDR.

Next we need to establish, that the sharing of incident data is necessary for the purpose of ensuring cyber security, and that there are no equally effective and less invasive methods available to achieve such legitimate interest. While it is difficult to measure the effect of sharing on security, most cyber security professionals and academics agree, that the sharing is the only effective method of protection against specific types of incidents [11, 23]. Irreplaceable role of sharing in cyber security is accepted also in EU legislation, the NIS directive states in recital 35 that "[providers] should be encouraged to pursue their own informal cooperation mechanisms to ensure the security of network and information systems. To encourage effectively the sharing of information and of best practice, it is essential to ensure that operators of essential services and digital service providers who participate in such exchanges are not disadvantaged as a result of their cooperation."

Establishment of provisional balance between controller's legitimate interest and rights and interests of data subject is a complex assessment, which should be conducted on case by case basis rather

[2] Confidentiality, integrity, availability.

than in an abstract manner. We can however derive specific criteria which should be taken into account. First we need to point out, that sharing of alert data benefits not only the peers that use sharing platforms, it also increases security of users of affected systems, some of which are at the same time data subjects, and wider community. Next we need to take into consideration the amount of data being shared. In some cases, some of the data are shared "just in case it is needed" and at the same time have very little influence on security. Another thing is that every controller should analyse severity of the incidents they are likely to prevent or mitigate and how widely those benefits can be shared with other peers. On the other side of the balancing test, where we deal with possible impact of sharing the data on data subject, we need to distinguish between different categories of data. Sharing of tools or information on malicious software will likely contain no or very little personal data. In the case of IoCs only very little data is being shared and even though it contains identifiers, it would be rather complicated for other peers to connect them to specific data subject. In these cases is the risk of processing for most data subjects very low. Far higher risk pose sharing of reports of security incidents and raw data, processing of which especially in combination with data shared by other peers may influence not only interest, but in some cases even fundamental rights of data subjects. So in some cases the controller should in the next step implement further measures in order to ensure higher protection of such data.

If after establishing a provisional balance the answer is not clear cut, one can go on to consider how introducing additional safeguards could prevent undue impact on the data subjects in order to help tip the balance in favour of the controller. These safeguards are discussed in following Section 5.

Last step is to demonstrate compliance and ensure transparency. Every controller should create proper documentation of the balance test and of any additional safeguards implemented. It is also a good idea to openly provide information about specifics of the processing not only to Data protection authorities but also to data subjects.

4.5 Other issues to consider

Another specific of the sharing platforms is the fact that some of the entities involved in its operation may have different position. For instance one entity might act as operator of the platform, while others are just nodes who share or receive shared data. This is from the legal point of view somewhat important, because there may be different obligations for each kind of entity. Especially for the operator of the platform, if the shared data are collected and further analysed, in operator's database. Due to discrepancies between operation of different platforms, it is necessary to always deal with legal position of each involved entity and legal basis for their mutual cooperation. In above mentioned situation for instance, the operator may be both in position of controller as well as processor of the personal data being shared.

Second issue is analysis of the data. In some sharing platforms the data is collected in one information system which conducts their analysis, and than distributes results of such analysis to nodes. Such analysis may in some cases fall within the following definition of profiling in the GDPR: "profiling means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's [...] reliability, [or] behaviour" [3]. If the system is for example creating reputation database of IP addresses based on collected data and IP's with bad reputation are then automatically blocked by nodes, it may cause violation of right of data subjects defined in Art. 22 of the GDPR. It states that "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal or similarly significant effects him or her" unless the decision is necessary for contract with the subject, authorised by public authority or based on subject's explicit consent.

Last issue we would like to stress is sharing of the data with nodes from outside the EU. Transfer of personal data outside the EU is in the GDPR forbidden unless adequate level of data protection is provided[3]. Transfers may be made only where the Commission has decided that a third country, a territory or one or more specific sectors in the third country, or an international organisation ensures an adequate level of protection, or if appropriate safeguards are in place. These safeguards may be provided for by:

- a legally binding agreement between public authorities or bodies;
- binding corporate rules (agreements governing transfers made between organisations within in a corporate group);
- standard data protection clauses in the form of template transfer clauses adopted by the Commission;
- standard data protection clauses in the form of template transfer clauses adopted by a supervisory authority and approved by the Commission;
- compliance with an approved code of conduct approved by a supervisory authority;
- certification under an approved certification mechanism as provided for in the GDPR;
- contractual clauses agreed authorised by the competent supervisory authority; or
- provisions inserted in to administrative arrangements between public authorities or bodies authorised by the competent supervisory authority[4].

So unless the territory in which the 'third-country node' resides is considered safe by the Commission or at least one of listed safeguards is in place, the sharing platform must be able to ensure complete anonymisation of data shared with this node or cease the sharing completely.

5 HOW TO MAKE THE SHARING LEGAL

In this section we discuss specific measures that controllers can and should implement in order to ensure compliance with EU data protection law and to prevent undue impact of sharing the data on the rights and interests of the data subject.

5.1 Privacy by design and default

One of the principles on which is based protection of personal data in the GDPR is protection by design and default. This principle is

[3] See Chapter V. of the GDPR

[4] See for instance online: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/transfer-of-data/

captured in the Art. 25 of the GDPR which states, that "the controller shall [...] implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing" and that "the controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed" [3].

The law does not provide any complete list of possible security measures, however GDPR specifically mentions some useful tools and refers to current state of the art. We also need to consider the fact that it is not necessary and in some cases even possible to always use all the tools. When determining which measures to implement, we have to take into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing.

In following subsections we propose basic technical, organisational and legal measures that can be used to ensure compliance with the GDPR. It is however not complete list, because first, there is too many possible tools, second, some platform require special attention and use of specific measures, and third, the state of the art is constantly changing and new tools may become available.

5.2 Technical measures

Technical measures of data protection in security alert sharing platforms can be divided in two main groups, security of storage and security of data transfer.

From the perspective of data storage security, there are only minor differences from storing any other personal information. The additional issue is the distribution of the data that can be spread across the sharing platform, especially in peer-to-peer platforms. Thus, all the nodes that store the data should be secured appropriately. However, it is often more reasonable to discard the data immediately after they are used due to their large volumes.

The problem of data distribution is closely tied to the second perspective of securing the alert sharing platform, the security of the data transfer. Securing the network connections via cryptographic means is nowadays a standard even in common network traffic. Proper authentication and authorization, using cryptographic means, is also needed to assure the legitimacy of peers, while increasing the overall trust in the sharing platform [20].

If a more detailed control over the distribution of the data is required, e. g., in a situation, where the piece of information can be shared with only a limited number of peers, additional features of the sharing platform should be implemented. Efforts were made to adapt well-known Traffic Light Protocol[5] (TLP) for the needs of automated information exchange. An example of such effort is the Information Exchange Policy (EIP)[6].

5.3 Organisational measures

Organisational measures can be divided into three main groups âĂŞ measures that relate to data, measures that relate to access to the data and measures related to organisation. First group include measures like data minimisation, limitation of storage period, second deals with access control and categorization of the data and the last periodical risk analysis and impact assessments.

Data minimisation is connected to the purpose limitation of processing of personal data mentioned above in the Section3.2. According to this principle, in the collection stage and in the following processing stage, personal data has to be fully avoided or minimised as much as possible. Consequently, personal data must be erased or effectively anonymised as soon as it is not anymore needed for the given purpose. This requirement applies in case of those who share data in the sharing platforms, as well as to its operator. Ability of the operator to control may be however relatively limited, due to the amount of shared data or technical specifics of the platform. One way how to deal with this issue is to require nodes to share data in structured form or in pre-defined categories, which would make the cleanup of the data easier.

The limitation of storage period is very similar. For example, some data about security incident may be relevant during ongoing attack or response (for example raw data for forensic analysis). This data however has often no further use for future analysis and therefore should be automatically deleted. In some cases would be also useful to implement regular cleanup procedure, which would analyse stored data delete those that are of no further use.

Another measure proposed directly in the GDPR is so-called Data protection impact assessment. If processing operations are likely to result in a high risk to the rights and freedoms of natural persons, the controller should carry-out a data protection impact assessment to evaluate, in particular, the origin, nature, particularity and severity of that risk. In some cases (systematic and extensive evaluation of personal aspects, processing of large scale data, systematic monitoring of publicly accessible area) it is even required prior to processing. Even though in our opinion the mere sharing of cyber security data itself does not lead to this obligation, in some cases[7] we would recommend to conduct at least simplified but documented assessment or risk analysis. The reason is that the outcome of the assessment might be extremely helpful when determining the appropriate measures to be taken in order to demonstrate that the processing of personal data complies with the GDPR.

Finally, last organisational measure that could be recommended is to regularly check if implemented safequards are sufficient and effective. The reason is that the quality, quantity or nature of collected or shared data might change, and that there could be new ways how to ensure compliance. Documented regular checks could also be used as a proof of compliance.

5.4 Legal measures

There are two main levels on which we can apply legal measures to ensure proper protection of personal data - the level of the platform and the level of individual nodes.

At the level of platform is most important to ensure that sufficient level of protection is ensured by each node. For this purpose, any operator of the platform should require each node to implement minimal security requirements which would define, what kind of data can be shared, how should be protected received data, and

[5] https://www.first.org/tlp

[6] https://www.first.org/iep/FIRST_IEP_framework_1_0.pdf

[7] For instance if the data controller is unsure which additional safeguards to implement.

what additional safeguards is each node required to implement. This can be done in the form of common rules for usage of the platform, or in the form of service-level agreement between the operator and each node.

At the level of individual nodes should be ensured protection of the data against unauthorised use. This could be done by implementing internal directives, which would define who can determine the methods and scope of data processing, who can access which data, how to handle them, and specify the responsibilities of individual employees. Another legal tool which should be always implemented are non disclosure agreements with every employee with access to the personal data.

6 CONCLUSIONS

Legal issues related to cyber security data sharing are extremely interesting and important research topic. This paper focuses particularly on personal data protection in security alert sharing platforms. Recent changes in the EU data protection law have prompted an intensive discussion about the extent to which personal data protection affects the use of these tools. This paper therefore summarizes main legal issues that should be considered when ensuring the compliance with the General data protection regulation and proposes tools and measures that can be used to deal with these issues.

First, we investigated what data are being shared using security alerts sharing platforms and for what purpose. Then we explained current development in data protection law in the EU and introduced its basic principles.

Second, based on these findings we analysed conditions for collection and sharing of personal data in these platforms, legal basis for their operation and identified main legal issues that need to be considered when ensuring compliance. Then we proposed legal, organisational and technical measures that can be implemented to ensure lawful operation of platforms and high level of protection of personal data.

The conclusions of this paper open issues that need to be addressed in the context of future research. In connection with the legitimate interest of the controller, it would be useful to conduct detailed analysis of how the balancing test should be conducted and which criterions should be taken into account. Other newly opened research question is what is the best legal solution that would allow non-EU organisations to participate in these sharing platforms.

ACKNOWLEDGMENTS

This research was supported by the Security Research Programme of the Czech Republic 2015 - 2020 (BV III / 1 VS) granted by the Ministry of the Interior of the Czech Republic under No. VI20162019029 The Sharing and analysis of security events in the Czech Republic.

REFERENCES

[1] Article 29 Working party. 2007. Opinion 4/2007 on the concept of personal data. (June 2007). http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

[2] Andrew Cormack. 2016. Incident Response: Protecting Individual Rights Under the General Data Protection Regulation. *SCRIPTed* 2016, 13:3 (2016), 258–282. DOI: http://dx.doi.org/10.2966/scrip.130316.258

[3] Council of European Union. 2014. Council regulation (EU) no 269/2014. (2014). http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1416170084502uri=CELEX:32014R0269.

[4] Court of Justice of the European Union. 2011. Judgement in Case C-70/10 Scarlet Extended SA v SociÃItÃI belge des auteurs, compositeurs et ÃIditeurs SCRL (SABAM). (Nov. 2011).

[5] Court of Justice of the European Union. 2016. Judgement in Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland,. (Oct. 2016).

[6] R. Danyliw. 2016. The Incident Object Description Exchange Format Version 2. RFC 7970 (Proposed Standard). (Nov. 2016). https://www.rfc-editor.org/rfc/rfc7970.txt

[7] H. Debar, D. Curry, and B. Feinstein. 2007. The Intrusion Detection Message Exchange Format (IDMEF). RFC 4765 (Experimental). (March 2007). http://www.ietf.org/rfc/rfc4765.txt

[8] ENISA. 2011. A flair for sharing - encouraging information exchange between CERTs. (Dec. 2011). https://www.enisa.europa.eu/publications/legal-information-sharing-1/at_download/fullReport

[9] ENISA. 2013. Detect, SHARE, Protect – Solutions for Improving Threat Data Exchange among CERTs. (Oct. 2013). https://www.enisa.europa.eu/activities/cert/support/information-sharing/detect-share-protect-solutions-for-improving-threat-data-exchange-among-certs/at_download/fullReport

[10] ENISA. 2014. Standards and tools for exchange and processing of actionable information. (Nov. 2014). https://www.enisa.europa.eu/publications/standards-and-tools-for-exchange-and-processing-of-actionable-information/at_download/fullReport

[11] Matthew H. Fleming and Eric Goldstein. 2012. Metrics for measuring the efficacy of critical-infrastructure-centric cybersecurity information sharing efforts. https://ssrn.com/abstract=2201033. (15 November 2012).

[12] Carol Fung and Raouf Boutaba. 2013. *Intrusion Detection Networks: A Key to Collaborative Security.* CRC Press.

[13] Graham Greenleaf. 2012. Global Data Privacy Laws: 89 Countries, and Accelerating. *Privacy Laws and Business International Report* 115, special issue (February 2012).

[14] Rick Hofstede, Pavel Čeleda, Brian Trammell, Idilio Drago, Ramin Sadre, Anna Sperotto, and Aiko Pras. 2014. Flow Monitoring Explained: From Packet Capture to Data Analysis with NetFlow and IPFIX. *Communications Surveys Tutorials, IEEE* 16, 4 (Fourthquarter 2014), 2037–2064.

[15] Guozhu Meng, Yang Liu, Jie Zhang, Alexander Pokluda, and Raouf Boutaba. 2015. Collaborative Security: A Survey and Taxonomy. *ACM Comput. Surv.* 48, 1, Article 1 (July 2015), 42 pages.

[16] Oscar Serrano, Luc Dandurand, and Sarah Brown. 2014. On the Design of a Cyber Security Data Sharing System. In *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security (WISCS '14).* ACM, New York, NY, USA, 61–69.

[17] Pavol Sokol, Jakub Míšek, and Martin Husák. 2017. Honeypots and honeynets: issues of privacy. *EURASIP Journal on Information Security* 2017, 1 (2017), 4.

[18] Jessica Steinberger, Anna Sperotto, Mario Golling, and Harald Baier. 2015. How to exchange security events? Overview and evaluation of formats and protocols. In *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM).* 261–269.

[19] Fredrik Valeur, Giovanni Vigna, Christopher Kruegel, and Richard A. Kemmerer. 2004. Comprehensive approach to intrusion detection alert correlation. *IEEE Transactions on Dependable and Secure Computing* 1, 3 (July 2004), 146–169.

[20] Tim van de Kamp, Andreas Peter, Maarten H. Everts, and Willem Jonker. 2016. Private Sharing of IOCs and Sightings. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security (WISCS '16).* ACM, New York, NY, USA, 35–38.

[21] Emmanouil Vasilomanolakis, Shankar Karuppayah, Max Mühlhäuser, and Mathias Fischer. 2015. Taxonomy and Survey of Collaborative Intrusion Detection. *ACM Comput. Surv.* 47, 4, Article 55 (May 2015), 33 pages.

[22] Cynthia Wagner, Alexandre Dulaunoy, Gérard Wagener, and Andras Iklody. 2016. MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security (WISCS '16).* ACM, New York, NY, USA, 49–56.

[23] W. Zhao and G. White. 2012. A collaborative information sharing framework for Community Cyber Security. In *2012 IEEE Conference on Technologies for Homeland Security (HST).* 457–462.