Cyber Crime Undermines Data Privacy Efforts – On the Balance Between Data Privacy and Security

Michael Mundt[1(✉)] and Harald Baier[2]

[1] Esri Deutschland GmbH, Bonn, Germany
m.mundt@esri.de
[2] Bundeswehr University - RI CODE Munich, Munich, Germany
harald.baier@unibw.de
https://www.esri.de, https://www.unibw.de/digfor

Abstract. The General Data Protection Regulation (GDPR) was put into effect in the European Union on 25th May 2018. GDPR aims to ensure the protection of personal data from individuals and the free movement of this personal data. Data privacy regulations are also currently being discussed nationwide in the United States of America and other countries. Regular guidelines of the European data protection board (edpb) support the technical GDPR implementation. However, cyber aggressors are increasingly succeeding in penetrating IT systems, e.g., by combining traditional ransomware techniques with data exfiltration. In this paper we address the trade-off between data protection as presumably regulated by the GDPR and the security implications of a hard and fast privacy enforcement. We argue that a too strict interpretation of the rules of data protection in the wrong place can even provoke the very reverse of data protection. The origin of our examination is to classify data in two GDPR relevant categories *personal data* (e.g., personal files of customers and company personal) and *IT operational data* (e.g. log files, IP addresses, NetFlow data), respectively. We then give a plea to strictly protect data of the first category and to handle the GDPR pragmatically with respect to the second one. To support our position we consider sample popular network protocols and show that it is low-threshold to exploit these protocols for data exfiltration, while the defender is only able to detect the attack on base of IT operational data. We hence emphasize the need for a new paradigm of risk assessment.

Keywords: Cyber Threat Intelligence · Data Breach · Regulatory Compliance · Insider Threat Management · Data Security and Privacy

1 Introduction

The objective of the General Data Protection Regulation (GDPR [19, Article 1]) is the protection of natural persons with regard to the processing of personal

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2023
Published by Springer Nature Switzerland AG 2023. All Rights Reserved
S. Goel et al. (Eds.): ICDF2C 2022, LNICST 508, pp. 417–434, 2023.
https://doi.org/10.1007/978-3-031-36574-4_25

data and rules relating to the free movement of data. However, today's cyber criminals stand in the way of these venerable goals, because they intend to smuggle out valuable data and then blackmail the victims by publishing stolen data or simply sell the data on one of the numerous trading platforms, e.g., on the darknet [6]. In order to ensure protection, the GDPR [19, Article 5] requires that processing is minimized to a necessary minimum and that worthy data is deleted immediately after processing. The data must only be processed for a specific purpose and, in general, the data must be stored in a form that allows the data subjects to only be identified as long as necessary for the processing purpose.

Personal data, that is in particular worthy for protection, is defined in a general manner [19, Article 4(1)]. It is defined as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". The question arises as to how the term "identifiable" is to be interpreted in concrete terms. Recital 26 attempts to remedy this: "[...] To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly". However, again the phrase "reasonably likely to be used" opens up room for interpretation when considering what is the data to focus protection measures on and when it comes to finding the right way to deal with data as it produced in all Open System Interconnection (OSI) [22] layers of an IT system such as log files and IP addresses.

In this paper we address the trade-off between data protection as presumably regulated by the GDPR and the security implications of a hard and fast privacy enforcement. We argue that a too strict interpretation of the rules of data protection in the wrong place can even provoke the very reverse of data protection. The origin of our examination is to classify data in two GDPR relevant categories *personal data* (e.g., personal files of customers and company personal) and *IT operational data* (e.g. log files, IP addresses, NetFlow data), respectively. While keeping the protection level of the first class of data, our main goal is to find a more balanced risk assessment view on the second data class. To the best of our knowledge such a *balance discussion* has not yet started.

Our use case to demonstrate the trade-off with respect to IT operational data is *data exfiltration*. We are first researching scientific dossiers on its threat landscape and contemporary technologies used for data exfiltration to adopt a first, simple classification scheme to categorize data exfiltration methods. In the next step, we identify relevant articles and considerations of the GDPR, which pursue the goal of protecting personal data. In addition, we evaluate the current guidelines of the European data protection board (edpb) [7] regarding the correct

behavior and notification in the event of a data breach. In doing so, we select relevant examples of the policy for our further investigation.

In order to discuss the balanced view on IT operational data, we consider sample widespread network protocols. As baseline of today's attack vectors we make use of the MITRE ATT&CK Framework [9] to exploit conventional network protocols. Our discussion incorporates a sample medium-sized company to determine which network protocols are mostly used on a normal work day. We show the low-threshold potential for misuse of the selected protocols by means of a proof of concept implementation and stress the importance of a more balanced privacy view on IT operational data to defend the attack.

Based on our sample use case we present our recommendations to implement an improved risk assessment for the protection of personal data and to take the security of processing data into account more consciously. The work pursues the concept of concentrating on the highly sensitive personal data and leaving the defender with the database that is necessary to thwart the considered attack vectors in good time.

The rest of the paper is organized as follows: Sect. 2 reviews related work in the scope of privacy and data exfiltration followed by a discussion of relevant articles of the GDPR for our work in Sect. 3. Next we present our sample use case and our practical discussion of the GDPR in the scope of 'data exfiltration' in Sect. 4 followed by our critical reflection of our results in Sect. 5. Section 6 concludes with the summary and an outlook for future efforts.

2 Related Work

In this section we review related work in the scope of our work. We first turn to related work in the context of GDPR in Sect. 2.1 followed by related work of IT operational data in our technical context of a data breach in Sect. 2.2.

2.1 Data Protection Related Work

A similar discussion of data categorization and a balanced view on data protection is sparse, however, a balancing consideration of privacy and security requirements due to Pope et al. [20] is available as preprint. The draft considers the counteracting requirements of privacy and accountability applied to identity management. Thus indicating the existence of an intrinsic tension between IT operations and data protection. Our work looks at this area of tension in greater detail and derives solutions for differentiating sensitive data in the sense of the General Data Protection Regulation during the operation of an IT process. Our work is much broader in scope, considering interactions with known attack vectors. We additionally consider external influences.

The notion of risks, as it is enshrined in the GDPR, is discussed in Gellert's paper [28]. He points out that in particular Art. 35 provides the obligation to carry out data protection impact assessments (DPIAs). However, this need is not yet linked to the attack vectors, i.e. actual cyber procedures. Our work focuses

on the needs arising from known cyber attack vectors. We derive a solution approach, based on a comprehensive risk analysis, to classify accruing data of the IT system as *IT operational data* and to treat it separately.

Furthermore some work in the scope of data protection implication assessment is available. For instance Bieker et al. [5] examines the new provisions of GDPR in detail and propose an adaptive process to suit the controller's needs. A balanced view on categories of personal data is not addressed by Bieker's process, though. We are introducing a more differentiated approach to the assessment of implications. Considering the interaction with cyber attack vectors yields important insights in the course of our work. Finally, we introduce the categories *personal data* and *IT operational data*. On this basis, we enable a more differentiated risk analysis and open up new possibilities for handling processed data without violating the requirements of the GDPR.

2.2 Data Exfiltration Related Work

Numerous considerations of technologies used for data exfiltration are available. A wide range of research has already been conducted in this domain in previous elaborations. Protective measures such as the simulation of current methods for data exfiltration are the subject of current work [18]. Here are a few more of the most recent publications. Covert drainage channels are very dangerous, as the victim often does not notice the process immediately. In this example, the exploitation of the properties of the Transmission Control Protocol (TCP) is examined in order to derive code of valuable data via the sequence number that can be deciphered on the receiver side [12]. Covert channels have also been detected between cloud instances in spite of existing countermeasures. The memory bus is exploited. The sender exists in the victim's environment as a trojan or any form of malicious program. The receiver exists in the attacker's environment. Both communicating entities execute without privileges. A rogue transmission channel is established [4, pp. 332,335]. Covert channels have also been demonstrated for Industry Control Systems (ICS). The basic idea is to log authentic ICS network data in normal operations for longer period and to alter this data afterwards with steganographic algorithm; valuable data is so hidden in network traffic by utilizing of various characteristics of network protocols [27]. Finally, a method is used here to exfiltrate sensitive data via the domain name service (DNS). Firewalls are typically configured to allow all packets on User Datagram Protocol (UDP) port 53. DNS is a mission critical service. Valuable data is encoded and hidden in the DNS query. DNS channels may be misused for stealing valuable and sensitive data [27]. It is not enough for us to prove only the abuse potential of individual technical protocols. We go two steps further in our work and put the misuse for data exfiltration into context with the goals of data protection and the currently known attack vectors of advanced cyber attackers. The technical misuse for data theft is part of an attack vector and this counteracts the goals of data protection. From this context, we derive concrete approaches to mitigate the risk for data protection.

Next, some papers are considered that evaluate different approaches to data exfiltration in a specific context. First, the Data Loss Prevention (DLP) System context is considered. Eight different technologies werde tested to determine whether relevant data could be diverted despite a DLP solution in place: Encrypting, compressing, changing file extension, renaming, splitting archives, deleting magic number of the files using Winhex (f.e. jpg starts with hexcode FF D8 FF and ends with FF D9) [15]. The next document tries to make a first classification of methods. For this purpose, a small selection of methods is examined. Three classes are proposed: content-based, header-based, meta-based. This is certainly not a complete classification, but it shows a way to prioritize and help IT experts to bring the issue of data exfiltration to the forefront of cybersecurity planning and actions within business [16, pp. 443,447]. Our work is not about classifying the various technical methods for data exfiltration. Rather, we are pursuing the path of finding a sensible classification for the data processed in an IT-system, so that a differentiated risk analysis and thus a carefully graded handling of the data is made possible.

Looking for more in-depth approaches to classification in the context of data exfiltration, the next paper offers an interesting approach. Data exfiltration countermeasures are classified in three major classes: preventive, detective, investigative. Furthermore, numerous countermeasures are assigned. Package inspection and anomaly based, for example are assigned to the detective class, each further subdivided into individual measures. Known channel inspection here is an example in the package inspection sub-class [1, Chapt. 5.1]. In addition, an attempt is also made here to classify the various methods of data exfiltration. This is more aggregated compared to the previous paper. The "Network" and "Physical" classes are executed and examples are provided for each [1, Chapt. 4.1]. In comparison, the MITRE ATT&CK Framework offers 9 techniques for data exfiltration within the enterprise matrix, i.e. a more differentiated subdivision in order to map them to the different attack vectors [9,10]. This paper concludes by recognizing that data exfiltration is a serious and ongoing issue in the field of information security and that the existence and emergence of such attack vectors make the countermeasures critical for an organisation's security [1, Chapt. 8]. In our work, we take up the quintessence of these works and design a new approach for a better differentiation of the risk and the inclusion of known cyber threats.

Our related work concludes with an extensive study of today's ransomware attacks. Here the trend is expressed that today's ransomware no longer blackmails the users with loss of access to data, but instead with a potential data leak of sensitive data [26, 197:7]. In addition, the ransomware is classified into three classes "detection", "defense", "prevention". This can be mapped to the previously used classification of methods for data exfiltration. This found the connection between ransomware and data exfiltration, which we use to compare the requirements of the GDPR and its corresponding guidelines [26, 197:11]. Our work is highly topical. The use of ransomware is currently a very big risk. Valuable data is being stolen more and more often in combination with ransomware, and we analyze exactly this danger with a view to the objectives of the GDPR.

Ultimately, our solution approach means potentially better protection for data and the freedom of natural persons within the meaning of the GDPR.

3 Relevant Articles of GDPR and Associated Guidelines

We first identify in Sect. 3.1 articles of the GDPR, which are relevant for our categorization of personal and IT operational data, respectively. We then review the edpb guidelines in Sect. 3.2 as a basis for our balanced risk assessment proposal.

3.1 Relevant Articles of GDPR

Arguably the most prominent article on data security in the GDPR is article 32 [19, Article 32]: security of processing. It states that suitable technical and organizational measures must be taken to ensure an appropriate level of protection. Therefore, state of the art technology, cost of implementation, nature, scope purpose of processing have to be taken into account. The risk of varying likelyhood and severity for the rights and freedoms of natural persons is incorporated into the design of the technical and organizational measures. Under the letters (a) to (d), methods such as encryption and pseudonymization as well as the protection goals of confidentiality, integrity, availability and resilience of processing systems and named services are listed. Furthermore, it then reads in article 32 2. that assessing the risk in particular the risk from accidental or unlawful destruction, loss, alteration and unauthorized disclosure shall be taken into account. The definition is directly related to the topic of data exfiltration that we are investigating. The process of data exfiltration results in the condition of unauthorized disclosure.

Recital 49 [19, Recital 49] further describes the measures to ensure security in processing by saying that personal data is processed to the extent strictly necessary and proportionate for the purposes of ensuring network and information security, i.e. the ability of a network or system to resist, at a given level of confidence, accidental events or unlawful or malicious actions that compromise the availability, authenticity, integrity and confidentiality of stored or transmitted personal data.

In principle, article 5 of the GDPR [19, Article 5] requires that personal (sensitive) data is only processed where the processing is necessary, is relevant to the attribution of the processing purpose and is lawful. Particularly interesting is also the requirement of the literal (e) that the identification of the data subject should only be possible for as long as it is necessary for the purposes for which it is processed. A tension is already emerging here. Exactly at the moment when data such as IP addresses or other traffic data are considered personal data, they are to be deleted after processing and are no longer available for analysis of the system or the services in long term. It will need to be considered how this affects data exfiltration countermeasures.

3.2 edpb Guidelines 2021/01 Version 2.0

Article 70 literal (d) [19, Article 70 (d)] enables the European data protection board (edpb) to issue guidelines in order to ensure the consistent application of the GDPR. Following this intention, these guidelines often provide examples explaining the correct implementation of the requirement stemming from the articles of the GDPR. To date, numerous guidelines have been issued. These are made available on the internet on the edpb websites[1]. We consider here the guidelines 01/2021 on examples regarding personal data breach notification. First, version 1.0 [7] of this guidance is considered. In this guideline, the definition of a data breach [19, Article 4 (12)] is first taken up. It is expressly pointed out that the consequences of data breach cannot be reversed perse and that preventive measures have to be taken in order to prevent a data breach [7, p. 6]. Examples are given in the following chapters 2–7. Some of these examples consider data breach due to accidental or intentional leaking of data. Chapter 3 specifically reports data exfiltration attacks.

Table 1. Examples regarding data breach notification with data exfiltration

Case No	Title
04	Ransomware without backup and with exfiltration
05	Exfiltration of job application data from a website
06	Exfiltration of hashed password from a website
07	Credential stuffing attack on a banking website
08	Exfiltration of business data by a former employee
17	Identity Theft
18	E-Mail exfiltration

Table 1 lists all sample cases regarding data exfiltration. We further explore the cases number 4 and number 18 in Table 1. The attacker has penetrated the victim system by abusing well-known protocols or functionality of the victim system. Advisable countermeasures are identified within each case study. For instance the edpb recommends for the case number 04 the forwarding or replicating of all logs to a central log server as one of these countermeasures. The importance of the availability of log files is already recognized at this point

4 Sample Attack Vectors for Balancing GDPR in Case of Data Exfiltration

This section shows for the sample use case of a potential data exfiltration in a small or medium sized company that IT operational data is essential to defend the data breach and hence to protect the actual personal data.

[1] https://edpb.europa.eu/our-work-tools/our-documents/publication-type/guidelines_de.

First, it's important to understand how a skilled attack is executed. For this purpose, we consider the MITRE ATT&CK Framework [9,10], in whose database the assessments of professional cyber analysts on worldwide incidents are incorporated. Our considerations focus on the enterprise matrix without distinguishing between individual platforms such as Windows, Linux containers, Cloud, etc. The attack vector for an enterprise is broken down into 14 phases (Fig. 1) that occur sequentially [11].

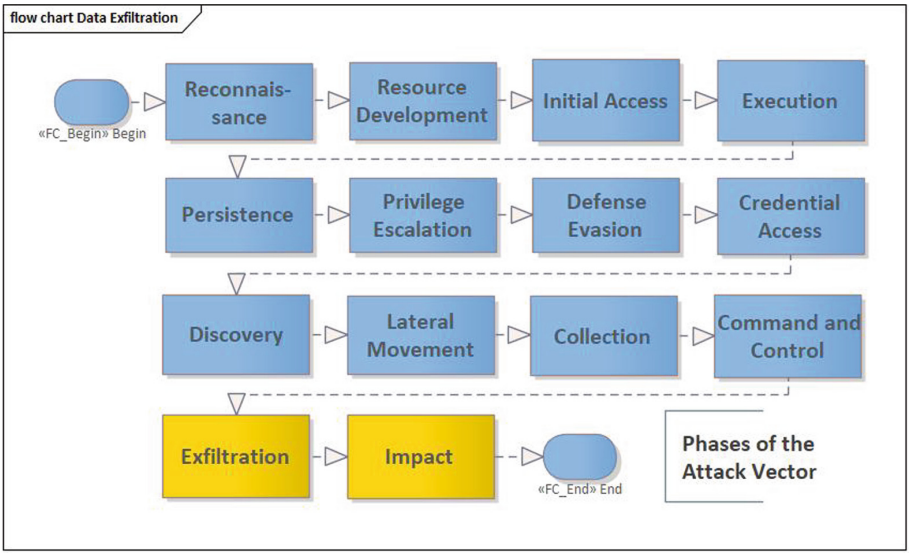

Fig. 1. MITRE ATT&CK Framework Matrix Enterprise

The attack vector is initiated by reconnaissance. Subsequently the attacker prepares himself in the resource development phase. The phase that is intended for initial access to the target system follows. At this point, the attacker is able to penetrate the victim's system. The next two phases, execution and persistence, are used to solidify access to the victim system. Now the attacker strengthen his position, gains rights on the system, disables defenses and gains access to credentials. The phases discovery, lateral movement and collection now serve to get an overview of valuable data and to collect them. Collecting can be understood, for example, as reading out data from information repositories. The attacker is now prepared. The final orders are issued via a command and control infrastructure, often Cobalt Strike, and then the data is exfiltrated. In the last phase, the effect is to be determined. If sensitive data is stolen, this will certainly have serious consequences.

We consider the phase Exfiltration in more detail. Here are two techniques of particular interest: 1) Exfiltration over alternative protocols 2) Exfiltration over web service. Item 1) indicates the abuse of protocols FTP, SMTP, https,

DNS, SMB. Item 2) indicates that websites are being exploited using SSL/TLS encrypted communication. Using the MITRE ATT&CK Framework, the connection between the exploitation of these protocols for data exfiltration and experienced groups (APT) can now be indicated, their attack targets tracked and the threat to one's own company better assessed.

In order to fully understand these possibilities, it is still important to understand the period over which data can be exfiltrated. Sensitive data can be partitioned and exfiltrated in small doses spread over time using one or more of the protocols mentioned. There are statistics on the length of time attackers spend unnoticed in the victim's IT system. Such a statistic for the years 2014 to 2019 can be seen in Fig. 2. The data for the last years 2020–2022 differ greatly.

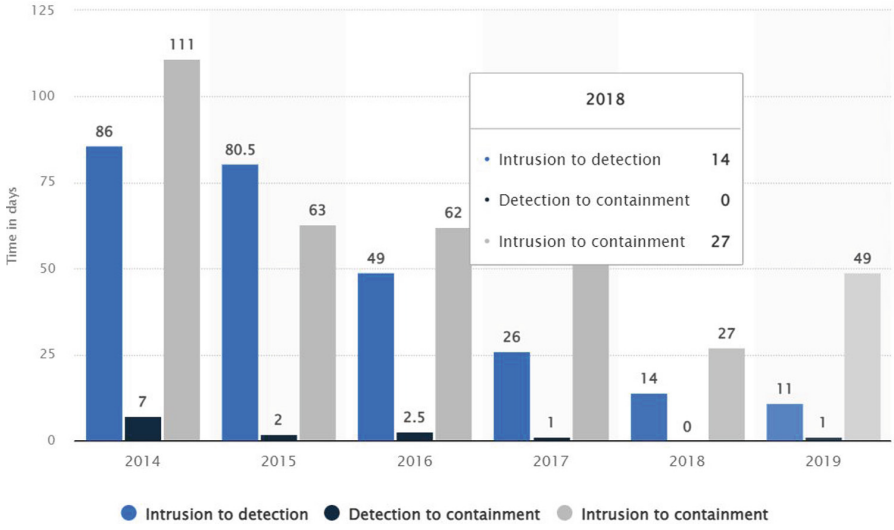

Fig. 2. Median time between intrusion and containment in industry [23]

The statistics were prepared for industrially used IT systems. These systems can be assumed to be well protected. Nevertheless, in 2019 a full 49 d passed before an intruder was detected and contained. Of course, not the entire period is used for dumping data. As previously discussed, the attack vector spans multiple phases. Nevertheless, it must be assumed that there is sufficient time to divert valuable data over days, weeks or even month. In less well-protected systems, the period of time before the attacker is discovered is likely to be much longer. Forbes is writing in 2021: "[...] Industry surveys over the years have shown dwell time ranging from a (sadly rare) best case of a couple of minutes to a worst case of hundreds of days. The average dwell time - depending on region, industry and who is generating the report - has varied widely" [17]. Mandiant, on the other hand, gives the global median dwell time with 21 d for 2022 much shorter then for the following years [21, p. 94].

4.1 Protocol Use Pattern Analysis

In order to get a better understanding of which protocols are used in a normal operation of a company, daily measurements are conducted at a medium-sized company. The results show in an aggregated form the 10 most commonly used network protocols for data transmission from within the company to outside or vice versa. The measurements are conducted on the firewalls of the gateway for the duration of one working day in the period from 10 a.m. to 4 p.m.

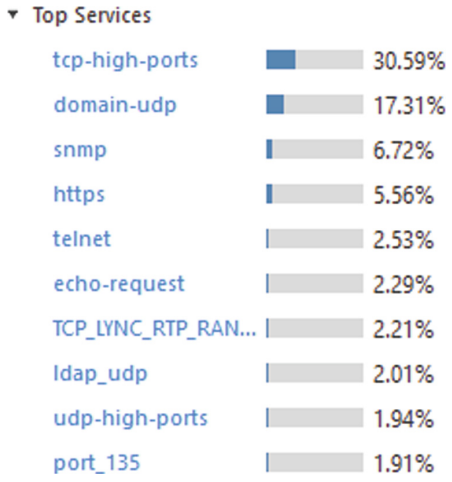

Fig. 3. Top ten ports used during a work day in an enterprise

It can be seen in Fig. 3 that TCP-based protocols in the port range of registered ports between 1014–49151 account for about a third. This is not surprising, since frequently used applications in the cloud, such as Drop Box and others, are located in this port compartment. This is followed by two port areas with domain-udp and snmp, which play an important role in the operation of an IT-network. The utility of the simple network management protocol (snmp) is that it allows information about networked devices to be collected across a variety of hardware and software types in a standardized way. IT-administrators very often utilize this protocol for network management purposes. The other protocol is most likely used to query the Domain Name Service (DNS). Finally the https protocol, which is used to browse websites, follows. Nowadays, websites are mostly accessed via the https protocol. The protocol https establishes a secure and encrypting connection by using the Transport Layer Security (TLS). Most often, the protocol is bound on TCP instead of UDP. The Fig. 3 shows the usage of the https protocol bound on TCP. The predecessor protocol http without encryption is hardly used anymore.

Figure 4 confirms this assessment. The Figure shows which protocols are mostly used by users. Again the top 10 protocols are displayed. The DNS service is requested most frequently, followed by the https protocol for accessing and

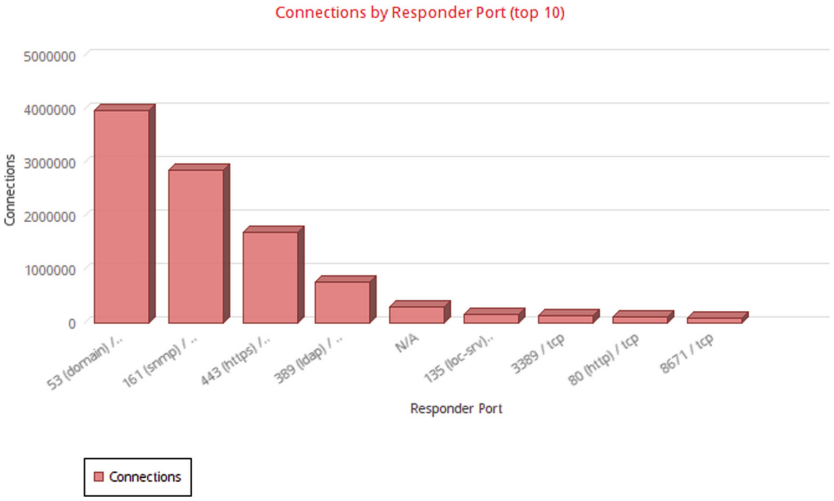

Fig. 4. Top ten services used during a work day in an enterprise

using websites. It's very understandable to see that DNS and https or, a few layer down within the OSI Model, TCP and UDP are used frequently. Most software applications in use today communicate with these protocols. From the perspective of an intruder who wants to exfiltrate data unnoticed, these commonly used protocols offer a good opportunity to hide individual data exfiltration messages in the noise floor.

4.2 Misuse of HTTP/2 Protocol as One Example of Many

The HTTP/2 protocol is the further development of the HTTP protocol. The Internet Engineering Task Force (IETF) [14] is specifying the HTTP/2 protocol. Therefore, the IETF applies the Request for Comments (RFC) procedure [2]. The new protocol brings performance benefits [13]. At the core of the new protocol is a binary framing layer. The HTTP messages are encapsulated in the protocol and then transmitted between client and server. We introduce important HTTP/2 terminology in Table 2.

HTTP/2 is only used for encrypted connections (Transport Layer Security (TLS) 1.2 or higher). The encryption method used employs digital certificates. A connection is established as follows: first of all, the protocol performs the TCP SYN-ACK handshake. Thereupon the TLS handshake follows.

Client and server exchange SSL certificates, cipher suite requirements and randomly generated data for creating session keys [8]. This procedure is shown in Fig. 5. Wireshark software was used to record the individual steps of the protocol. The individual steps are visualized by the software used. In addition, Wireshark offers the functionality to search through each individual entry afterwards in deep detail, up to the hex code analysis. These steps mark the connection

Table 2. HTTP/2 important terminology

Term	Description
Stream	Sequence of frames mapping to a logical request or response message
Frame	Smallest unit in HTTP/2 communication, each containing a frame header which identifies the stream to which the frame belongs to
Message	A complete sequence of frames that map to a logical request or response message

setup between transmitter and receiver. First, the prerequisites for encrypted transmission are created. Having the connection established, client and server exchange binary frames. All frames consist of a common 9-byte header (length of the frame), type, a bit field for flags, and a 31-bit stream identifier. Additionally, a 24-bit length field allows each single frame to carry up bytes of data as frame payload. The protocol is capable of multiplexing. Multiple streams can be transmitted over one connection. For each stream, the HEADERS frame (Fig. 6) is transmitted first, followed by the DATA frames. Various scripting and programming languages provide libraries to use the features and functions of the HTTP/2 protocol. The interpreted, higher-level programming language Python provides various libraries with an implementation of the protocol.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000000	192.168.179.45	192.168.179.1	DNS	71	Standard query 0xae1f A www.bing.de
2	0.000357108	192.168.179.45	192.168.179.1	DNS	71	Standard query 0xbd21 AAAA www.bing.de
3	0.022686114	192.168.179.1	192.168.179.45	DNS	87	Standard query response 0xae1f A www.bing.de A 204.79.197.219
4	0.025030755	192.168.179.1	192.168.179.45	DNS	145	Standard query response 0xbd21 AAAA www.bing.de SOA ns1-204.a…
5	0.025777536	192.168.179.45	204.79.197.219	TCP	74	44812 → 443 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM=1 …
6	0.049649365	204.79.197.219	192.168.179.45	TCP	66	443 → 44812 [SYN, ACK] Seq=0 Ack=1 Win=65535 Len=0 MSS=1440 W…
7	0.049748838	192.168.179.45	204.79.197.219	TCP	54	44812 → 443 [ACK] Seq=1 Ack=1 Win=64256 Len=0
8	0.119701475	192.168.179.45	204.79.197.219	TLSv1.2	571	Client Hello
9	0.144185430	204.79.197.219	192.168.179.45	TCP	56	443 → 44812 [ACK] Seq=1 Ack=518 Win=525056 Len=0
10	0.147662974	204.79.197.219	192.168.179.45	TCP	1506	443 → 44812 [ACK] Seq=1 Ack=518 Win=525056 Len=1452 [TCP segm…
11	0.147716829	192.168.179.45	204.79.197.219	TCP	54	44812 → 443 [ACK] Seq=518 Ack=1453 Win=64128 Len=0
12	0.147779477	204.79.197.219	192.168.179.45	TCP	1506	443 → 44812 [ACK] Seq=1453 Ack=518 Win=525056 Len=1452 [TCP s…
13	0.147810405	192.168.179.45	204.79.197.219	TCP	54	44812 → 443 [ACK] Seq=518 Ack=2905 Win=62848 Len=0
14	0.147839918	204.79.197.219	192.168.179.45	TCP	1506	443 → 44812 [ACK] Seq=2905 Ack=518 Win=525056 Len=1452 [TCP s…
15	0.147859867	204.79.197.219	192.168.179.45	TLSv1.2	404	Server Hello, Certificate, Server Key Exchange, Server Hello …
16	0.147870330	192.168.179.45	204.79.197.219	TCP	54	44812 → 443 [ACK] Seq=518 Ack=4357 Win=61440 Len=0
17	0.147893580	192.168.179.45	204.79.197.219	TCP	54	44812 → 443 [ACK] Seq=518 Ack=4707 Win=61184 Len=0
18	0.151961165	192.168.179.45	204.79.197.219	TLSv1.2	212	Client Key Exchange, Change Cipher Spec, Encrypted Handshake …
19	0.178931623	204.79.197.219	192.168.179.45	TCP	56	443 → 44812 [ACK] Seq=4707 Ack=676 Win=524800 Len=0
20	0.181393200	204.79.197.219	192.168.179.45	TLSv1.2	380	New Session Ticket, Change Cipher Spec, Encrypted Handshake M…
21	0.181448268	192.168.179.45	204.79.197.219	TCP	54	44812 → 443 [ACK] Seq=676 Ack=5033 Win=64128 Len=0
22	0.182273188	192.168.179.45	204.79.197.219	TLSv1.2	297	Application Data
23	0.182986493	192.168.179.45	204.79.197.219	TLSv1.2	259	Application Data

Fig. 5. Wireshark recording from using HTTP/2 protocol

For this we use the "hyper" library [3] to implement an HTTP/2 client. This is just one library among many others that can be used to execute the HTTP/2 protocol. We chose this library because we want to use the Python scripting language and the implementation is comparatively easy using this library. The HTTP/2 client integrates into the Python requests library. The entire connection establishment is carried out via mounting of the HTTP20Adapter to the current session. The data is then send to the server using the POST method of

the HTTP/2 protocol. In this case it is not necessary to encrypt the data for exfiltration. This happens anyway via the HTTPS connection between client and server. The defender of the IT victim system cannot inspect this communication due to the encryption.

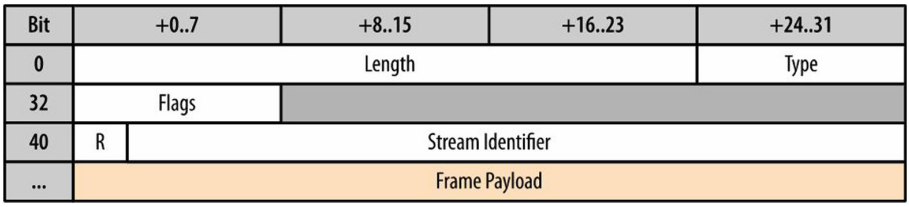

Fig. 6. Nine byte frame header [13]

The wireshark recording in Fig. 7 clearly shows the encrypted transmission of data over the HTTP/2 - HTTPS connection. This transfer does not require higher privileges on the victims system. In this manner we can show that valuable data may be exfiltrated using the HTTP/2 protocol. The excerpts of the Wireshark recording are suitable for deepening our solution. First, we have seen the establishment of the connection and the negotiation of the encryption. We assign this data to the category *IT operational data*. This data contains potentially Personal Identifiable Information (PII) as defined by GDPR. IP and MAC addresses are included. These are declared as PII. More data to come: the connection is established via a process in the IT-system that is executed under a user ID. Data such as the call of the process, the user ID, the time of execution, etc. are inevitably recorded in the IT system. In addition, there are the calls to system functions such as reading data files. This data is also written to log files, mostly managed automatically by the IT system. Sophisticated attackers might use running processes, induce a thread, allocate memory, and copy shellcode into it, which then performs the data exfiltration. Every action leaves digital traces (data) in the IT system. The Wireshark extract offers a first insight into this fact (Fig. 5).

We recommend categorizing this data as *IT operational data* and valuing their particular value for detecting cyber attack vectors in the risk assessment. To explain in more depth the value of this data for detecting attacks, we make an example using the MITRE ATT&CK framework. We pick up a Technique noted in it: Scheduled Transfer[2]. This Technique is used to achieve data exfiltration. To be able to detect the use of this Technique, among other things, the monitoring of network traffic flow[3] and monitoring the creation of network connections[4]

[2] https://attack.mitre.org/techniques/T1029/.
[3] https://attack.mitre.org/datasources/DS0029/#Network%20Traffic%20Flow.
[4] https://attack.mitre.org/datasources/DS0029/#Network%20Connection%20Creation.

are proposed. Precisely at this point our approach is strengthened. If the *IT operational data* is deleted too early, this Technique and thus the attack vector can no longer be detected and a complete clarification in retrospect is denied.

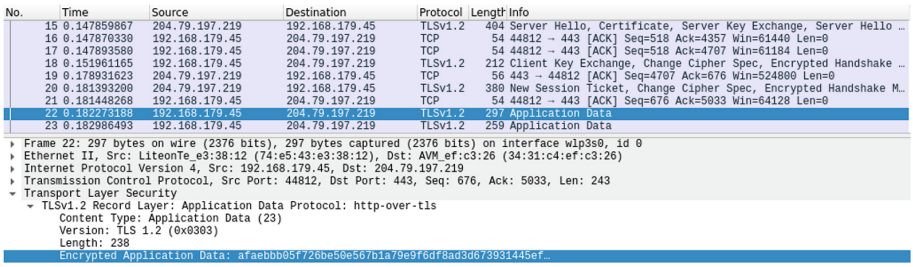

Fig. 7. Wireshark recording from encrypted sending of data

Also, the original sensitive data about persons (PII) are processed and are reflected in the Wireshark extract (Fig. 7). At this point, however, this data is already encrypted. This data was previously read from files and encrypted by a process. We recommend that you categorize this original data as *personal data*, protect it as much as possible and delete it as soon as possible. Our categorization in *personal data* and *IT operational data* allows to draw different conclusions in the risk analysis and to manage the data differently without lowering the level of protection of personal data. On the contrary, the level of protection is increased. If PII is stolen and encrypted by cyber-attackers, the guarantee of protection and free flow of information as required by the GDPR can no longer be assured with own strength. Our approach helps to avoid this helpless situation.

4.3 Malicious Reconfiguration of an E-Mail Client

Selected case No. 18 deals with the exfiltration of data via e-mails. Unlike in case No 4 no technical protocols are exploited here. Instead, it is described how e-mail clients used were covertly reconfigured. The configuration happens in such a way that all messages are filtered according to certain criteria and then automatically forwarded to an external e-mail address including any attachments. The attacker can easily acquire knowledge of the configuration via the public support pages of the respective manufacturers [25]. In addition, documentation for all commercial products is openly available online [24]. After the privilege escalation phase of the attack vector, but at latest after the evasion of the defense measures and access to credentials, the attacker is able to carry out these configurations [11]. This procedure can be imagined especially in the case of an insider. The re-configuration of the e-mail client usually requires higher privileges on the victim's IT system.

In order to be able to track these changes to the configuration afterwards, it is necessary to keep a log file in which each change is saved. Ideally, the system

monitors configuration changes of the account in a way, that malicious activities are recognized immediately or at least in case of an investigation. Therefore, the system logs changes of account configuration in date and time. We recommend providing this information to the defense attorney. In order that the information is not deleted immediately after too short a period, we recommend that this data be categorized as *IT operational data*.

5 Balancing GDPR Requirements Against the Hazards

In Sect. 4 we show the exploitation of common IT network protocols for data exfiltration by misusing their technical possibilities. The criminal exfiltration of personal data obviously contradicts the protection goal of the integrity of this data, as formulated in article 5 [19, Article 5 Chap. 1.f.], in article 32 [19, Article 32 Chap. 1.b.], as well as the superior goal of protecting natural persons and their personal data [19, Article 1]. The selection, implementation and regular review of the technical and organizational measures must be checked accordingly in order to contain the threat.

In doing so, it is not sufficient to look isolated at the risk of data exfiltration. The GDPR requires careful risk assessment to ensure a level of security appropriate to the risk [19, Article 32 Chap. 1.]. The entire attack vector must be included in the risk assessment and hence all attack vector tactics must be included in the analysis to meet the threats. The analysis must be performed backwards and forward-looking, over the entire time which might even be longer than a year. Sometimes data theft only becomes known when the personal data concerned is sold on marketplaces on the darknet or confiscated by law enforcement. If the *IT operational data* is already deleted, there will be no way to clarify the incident. The vulnerability used at the time remains undetected.

In the different phases of the attack, digital traces are left behind, which can be reflected in log files, among other data sources. The use of network protocols must be checked at least on a random basis (flow data). Log usage times and patterns need to be analysed for suspicious traces. Log data must be examined to determine whether a mechanical, automated attack on e.g. firewalls or internal services has taken place. In the "Defense Evasion" tactic, the attacker tries to switch off protective measures such as a virus scanner. Here, too, traces are left behind which must now be found as evidence. In the Tactic "Collection" data is collected. Database requests, copying operations or downloads from company core services leave traces here. Repeated access to data sources in the company's core systems should be monitored and critically questioned. Otherwise, it is very likely that the use of these protocols will only be determined retrospectively when the high-quality data has already been dumped.

This has an impact on the consideration of the requirement in article 5 [19, Article 5 Chap. 1.e.] which allows personal data to be kept for as long as it is for statistical purposes in accordance with article 89(1) as subject to implementation of the appropriate technical and organisational measures required by GDPR in order to safeguard the rights and freedoms of the data subject ('storage limitation'). Storage limitations must be carefully considered and not be designed

to be shorter than the full duration of the attack vectors. Better longer. This consideration affects data of our category *IT operational data* such as logfiles and IP-addresses within. Early deletion or too excessive anonymisation of data in this category may jeopardise the full assessment of the risk to which the personal data is or has been exposed during processing. It becomes clear that the category of *IT operational data* is essential to combat the identified threats vectors. The data of this category must be available to the defender in sufficient time and detail, even though there is personal data within. Deletion immediately after processing is no longer an option due to the risks. The situation is different with the first category of *personal data*. Anonymization during processing and deletion after processing must be implemented as quickly as possible. Concentrating on this means sustainable data protection through the beneficial interaction with data security.

6 Conclusion and Future Work

In order to achieve the goals of the GDPR a risk assessment must take place for the processing of personal data. The dangers emanating from an attack vector of experienced cyber attackers must be taken into account. The defender must be able to store, to analyze and to document all the *IT operational data* necessary to evaluate traces across all phases of the attack vector and thus recognize intruders as best as possible or even to investigate the case of data theft retroactively. Without sufficient data for evaluation by the defender, a successful attack may result and the exfiltration of important data cannot be prevented in time. At this point, we recommend to focus on the protection of valuable *personal data* and to make the *IT operational data* available to the defense in a way assisting the defense counsel and in this way, to promote the complete clarification and traceability of digital forensics.

In future work, we will drill deeper the investigation on misusing network protocols for criminal data theft. In particular, the cookie technology is disassembled with intent to abuse. In addition, we will investigate how knowledge of the potential for misuse can be used for a simulation in order to gain insights into the effectiveness of protective and mitigating measures.

References

1. Ullah, F., Edwards, M., Ramdhany, R., Chitchyan, R., Babar, M.A., Rashid, A.: Data exfiltration: a review of external attack vectors and countermeasures. J. Netw. Comput. Appl. **101**(2), 18–54 (2017). https://eprints.lancs.ac.uk/id/eprint/88549/1/1_s2.0_S1084804517303569_main.pdf
2. Belshe, M., Peon, R., Thomson, M.: Hypertext transfer protocol version 2 (HTTP/2) (2015). https://datatracker.ietf.org/doc/html/rfc7540. Accessed 07 Mar 2021
3. Cory Benfield. Hyper: HTTP/2 client for python (2015). https://hyper.readthedocs.io/en/latest/. Accessed 13 Mar 2022

4. Semal, B., Markantonakis, K., Mayes, K., Kalbantner, J.: One covert channel to rule them all: a practical approach to data exfiltration in the cloud. In: 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) TRUSTCOM Trust, pp. 328–336 (2020)
5. Bieker, F., Friedewald, M., Hansen, M., Obersteller, H., Rost, M.: A process for data protection impact assessment under the European general data protection regulation. In: Schiffner, S., Serna, J., Ikonomou, D., Rannenberg, K. (eds.) APF 2016. LNCS, vol. 9857, pp. 21–37. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44760-5_2
6. Darktrace Blog and Dianna Leddy. Double extortion-ransomware (2021). https://www.darktrace.com/de/blog/double-extortion-ransomware/?utm_source=xing&utm_medium=static-awareness-de&utm_campaign=campaign_socialmedia&dclid=CMnvw4O-2vICFdJD4AodzLAPWw. Accessed 22 Oct 2021
7. European Data Protection Board. Guidelines 01/2021 on examples regarding data breach notification, version 2.0 (2021). https://edpb.europa.eu/system/files/2022-01/edpb_guidelines_012021_pdbnotification_adopted_en.pdf. Accessed 06 Mar 2022
8. Cloudflare. What happens in a TLS handshake? — SSL handshake (2022). https://www.cloudflare.com/learning/ssl/what-happens-in-a-tls-handshake/. Accessed 13 Mar 2021
9. MITRE Corporation. MITRE ATT&CK framework (2021). https://attack.mitre.org/. Accessed 04 Mar 2021
10. MITRE Corporation. MITRE ATT&CK navigator (2021). https://mitre-attack.github.io/attack-navigator/. Accessed 04 Mar 2021
11. MITRE Corporation. MITRE ATT&CK navigator - matrix enterprise (2022). https://attack.mitre.org/matrices/enterprise/. Accessed 08 Mar 2022
12. Goverman, J., Tekeoglu, A.: Stealthy data exfiltration via TCP sequence numbers based covert channel. In: 2021 International Conference on Computer Information and Telecommunication Systems, 1–5 Nov 2021. https://ieeexplore.ieee.org/document/9618137
13. Gregorik, I.: High performance browser networking HTTP/2 (2013). https://hpbn.co/http2/. Accessed 13 Mar 2021
14. IETF HTTP Working Group. Http/2 (2015). https://http2.github.io/. Accessed 13 Mar 2022
15. AlKilani, H., Nasereddin, M., Hadi, A., Tedmori, S.: Data exfiltration techniques and data loss prevention system. In: 2019 International Arab Conference on Information Technology (ACIT) Information Technology (ACIT), pp. 124–127 (2019)
16. King, J., Bendiab, G., Savage, N., Shiaeles, S.: Data exfiltration: methods and detection countermeasures. In: 2021 IEEE International Conference on Cyber Security and Resilience (CSR) Cyber Security and Resilience (CSR), pp. 442–447 (2021). https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9527962
17. Saryu N.: Why the dwell time of cyberattacks has not changed (2021). https://www.forbes.com/sites/forbestechcouncil/2021/05/03/why-the-dwell-time-of-cyberattacks-has-not-changed/?sh=48b387a457d8. Accessed 06 Nov 2022
18. Mundt, M., Baier, H.: Towards mitigation of data exfiltration techniques using the MITRE ATT&CK framework. In: 12th EAI International Conference on Digital Forensics & Cyber Crime (EAI ICDF2C). https://compass.eai.eu/events/detail/242/eai-icdf2c-2021
19. European Parliament. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation) (2016). https://eur-

lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679. Accessed 06 Mar 2022
20. Pope, N., Goodell, G.: Identification for accountability vs privacy (2022). https://arxiv.org/ftp/arxiv/papers/2201/2201.06971.pdf. Accessed 01 Apr 2022
21. Mandiant Special Report. M-trends 2022 (2022). https://www.mandiant.com/media/15671. Accessed 06 Nov 2022
22. Salvi, M.V., Bapat, M.P.: Mode of data flow in the OSI model. IJIERT - Int. J. Innov. Eng. Res. Technol. **2**(3), 1–7 (2015)
23. Statista. Median time period between intrusion, detection, and containment of industrial cyber attacks worldwide from 2014 to 2019 (2020). https://www.statista.com/statistics/221406/time-between-initial-compromise-and-discovery-of-larger-organizations/. Accessed 07 Mar 2021
24. Microsoft Support. Configure email forwarding in Microsoft 365 (2022). https://docs.microsoft.com/en-us/microsoft-365/admin/email/configure-email-forwarding?view=o365-worldwide. Accessed 11 Mar 2022
25. Microsoft Support. Use rules to automatically forward messages (2022). https://support.microsoft.com/en-us/office/use-rules-to-automatically-forward-messages-45aa9664-4911-4f96-9663-ece42816d746. Accessed 11 Mar 2022
26. McIntosh, T., Kayes, A.S.M., Chen, Y.P.P., Ng, A., Watters, P.: Ransomware mitigation in the modern Era: a comprehensive review, research challenges, and future directions. ACM Comput. Surv. (CSUR). **54**(9), 1–36. ACM, New York, NY (2021)
27. Neubert, T., Vielhauer, C., Kraetzer, C.: Artificial steganographic network data generation concept and evaluation of detection approaches to secure industrial control systems against steganographic attacks. In: The 16th International Conference on Availability, Reliability and Security, pp. 1–9 (2021). https://doi.org/10.1145/3465481.3470073
28. Gellert, R.: Understanding the notion of risk in the general data protection regulation (2016). https://www.sciencedirect.com/science/article/abs/pii/S0267364917302698. Accessed 09 Apr 2022