DATA AND APPLICATION SECURITY

LECTURE NOTES

Pavel Laskov

December 16, 2019

Contents

Chapter 1

Current Threat Landscape

Cybersecurity plays a crucial role in the modern world economy. According to Gartner's forecast [11], investment in cybersecurity is expected to reach 125 billion US$ in 2019 and to grow at 8-12% rate until 2022. "Security leaders are striving to help their organizations securely use technology platforms to become more competitive and drive growth for the business," said Siddharth Deshpande, research director at Gartner. The prime goal of security investment is to minimize technical and business losses arising from security incidents. Due to the high complexity of modern information systems the risk of security incidents cannot be eliminated, yet various countermeasures enable organizations to minimize potential losses and hence optimize their operations.

Several factors have a crucial impact on security investment. Digital transformation of many traditional businesses broadens the target base for cyberattacks. It also exacerbates the lack of expertise in the field of cybersecurity since a larger number of organizations have to address their exposure to security risks. Regulatory changes like the EU's General Data Protection Regulation (GDPR) are also substantial drivers for security investment. The need for understanding and implementation of the new regulatory provisions has spurred a demand for legal as well as technical consulting services in the field of data privacy.

Key trends in the security investment in 2017-2019 are summarized in Figure 1. While substantial resources are still invested in traditional on-premise solutions, such as network security equipment, identity and access management as well as infrastructure protection, a clear trend is the increasing role of security services. The growing importance of security services reflects the sustained shortage of security expertise in the field as well as the increasing specialization of security businesses. Integrated risk management remains in high demand, which underlines the persistence of security risks and the importance of effective risk management instruments. The latter enable organisations to attain a holistic view of their security risks, develop effective minimization strategies and define metrics for measuring the success of security operations.

The strongest growing area in security investment is cloud security (60-70%). While the absolute investment volume in cloud security is still substantially lower than in other areas presented in Figure 1 (up to 500 million US$), such a strong growth rate reflects the rapid migration of IT infrastructure into cloud and the respective demand for cloud-specific security solutions.

The shortage of qualified security professionals and the resulting high demand and cost of security consulting drive the demand for security automation tools. A recent study of IT security staffing trends [10] showed that 29% of respondents already deploy security automation

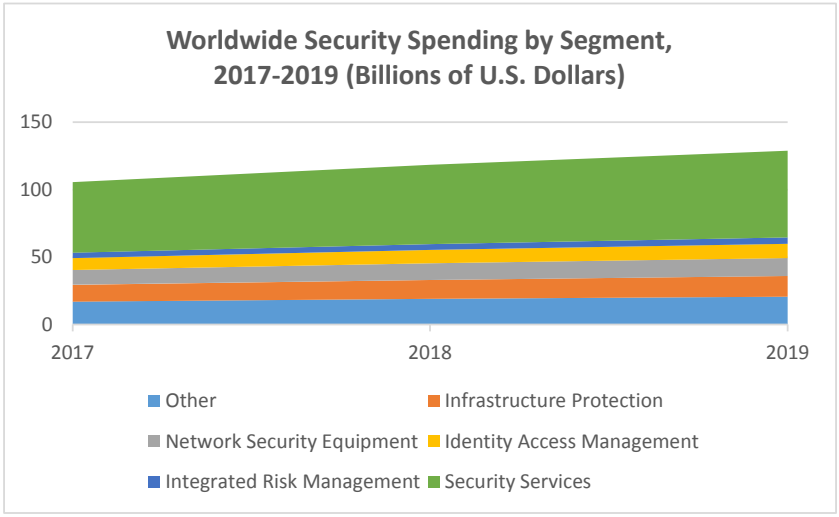

Figure 1: Key areas in security investment (2017-2019).

and another 50% are planning to roll out security automation in the next three years. The main benefits of security automation lies in reduction of manual effort on routine security tasks, such as identification and patching of known vulnerabilities, management of certificates and access control credentials, keeping track of hardware and software assets. All of these tasks inevitably lead to collection and processing of large amounts of data and hence necessitate the deployment of data analysis tools, such as AI and machine learning.

1.1 Losses from Security Incidents

Investment in IT security can only be justified if it helps to eliminate losses from security incidents. Understanding and quantification of such losses is an indispensable part of security risk analysis. In this section we present a methodology for assessment of losses for one specific type of security incidents – data breaches – and discuss some findings from application of this methodology.

Data breach is defined as a security incident in which certain data records are stolen. Typically data breaches involve personal records of customers such as addresses, contact data, bank account data, credit card numbers, medical records, to name only a few. Stolen customer data can be used by attackers directly for gaining profits, e.g., by selling credit card data on black markets, or as a means for further attacks, e.g., extortion. Some of the data lost in data breaches can be extremely sensitive. For example, the data stolen in the incident at the US Office of Personnel Management included records of lie detector tests for persons seeking advanced government clearances and containing intimate details of their personal life. As a result of that incident, the US Government is believed to have spent more than 1 billion US$ in victim compensations and follow-up services, such as credit score monitoring and identity theft insurances.

The costs of data breaches were analyzed in a series of annual studies conducted by the

Ponemon Institute (e.g., [6] for 2018). Such studies addressed the following goals:

- Estimate the absolute and the relative cost of a data breach.
- Understand the cost structure for data breaches.
- Analyze the weight of various cost factors for data breaches in different industries.
- Analyze the weight of various root causes of data breaches.
- Analyze the impact of different influencing factors.

The data for the study was gathered by recruiting 447 organizations that suffered data breaches and interviewing more than 2200 individuals in these organizations who were knowledgeable about these incidents. As a basic information, the respondents were asked to provide the number of records lost in a data breach and to estimate the percentage of customer base lost as a result of an incident. In the course of follow-up interviews more details about the incident costs were revealed. The data collected in this way reflects the spending on the discovery of and the immediate response to the data breach, e.g., forensics and investigations, as well as the follow-up costs, e.g., victim notification and compensation, legal and compliance costs, expenditures on re-gaining customers, etc. Additional details discussed in the interviews comprised the presumed root causes of incidents as well as different operational factors that may have had a positive or a negative impact on the incident costs.

For the systematic analysis of the cost structure, the costs were classified into the following categories:

- *Immediate costs:* conducting investigation and forensics, determining the victims, incident response and containment, preparation of disclosure, communication and PR activities.
- *Aftermath costs:* audit and consulting, legal costs, victim compensation, lost business, renewed customer acquisition costs.

From the operational perspective, the ascertained costs are further classified into the following categories:

- *Direct costs:* expenses paid out for certain activities.
- *Indirect costs:* time, effort and resources spent on incident handling.
- *Opportunity costs: estimates of lost business due to reputation loss, customer churn, etc.*

In the course of subsequent analysis the average cost of the data breach as well as the average cost of the stolen data record was computed and profiled along several dimensions: country, industry, company size, root cause, as well as with relation to additional influencing factors.

The average cost of a data breach incident in 2018 was $3.86 million, which is a 6.6% increase compared to $3.62 million in 2017. The highest average cost of a data breach was observed in the US and the Middle East countries ($7.91 million and $5.31 million, respectively);

the lowest average data breach cost was observed in Brasil and India ($1.24 million and $1.54 million, respectively). These results clearly show that the cost of security incidents is closely correlated with the labor cost of highly skilled IT specialists in the respective countries. The average cost per lost data record across all companies and regions was $148.

The distribution of data breach costs per industry demonstrates the affinity of the costs with the intuitive "value" of the lost data. The highest costs per record were observed in the healthcare and financial industries ($408 and $206, respectively) which are heavily regulated. The lowest costs per record were observed in public organizations and research ($75 and $92, respectively) which are much less profit-oriented and hence experience lower aftermath costs related to customer loss.

Among the root causes for data breaches, malicious hacker activities are most common (48%) followed by a human error (27%) or a technical glitch 25%). Likewise, data breaches caused by malicious attacks are more costly ($157 per record) than the other two ($128 and $131, respectively). The geographical distribution of data breach root causes varies substantially. The Middle East countries are the strongest hit by malicious attacks (61% of all data breaches) whereas Turkey exhibits the lowest ratio of malicious activity (38%) and the highest ratio of technical glitches (33%) as data breach root causes. Italy has the highest ratio of a human error as a data breach root cause (35%).

Several factors affect the cost of data breaches. The most prominent mitigating factors were the existence of an incident response team (-$14.0) and extensive use of encryption (-$13.1). These findings are not surprising. Other important mitigating factors are existence of business continuity modeling (-$9.3), employee training (-$9.3), participation in threat sharing (-$8.7) and deployment of artificial intelligence security tools (-$7.2). On the other side of the influencing factor spectrum lie the involvement of third parties in a security breach ($13.4), extensive cloud migration ($11.9), compliance failures ($11.9) and extensive use of mobile platforms ($10.0). These findings demonstrate that regulation of security related aspects does help to decrease the data breach costs, whereas unusual technical platforms such as cloud or mobile devices make incident discovery and response more difficult and adversely affect the data breach costs.

A different type of analysis is needed to understand the costs of "mega-breaches" involving the loss of more than 1 million records. For this study, 11 companies that experienced such breaches in 2018 were recruited and their various costs were recorded. Since the sample of 11 companies is too small for statistical analysis, Monte-Carlo simulations were carried out to estimate the most likely cost structure of mega-breaches. Monte-Carlo simulation uses repeated sampling from distributions with parameters determined by empirical data. The means and the standard deviations of the predicted outputs enable one to assess the most likely outcomes of the target values. The simulated total cost of mega-breaches is shown in Figure 2. The simulation reveals that the average total cost of a mega-breach is expected to be up to $350 million for a breach with a loss of 50 million records. Likely outcomes for the cost of a 50 million record breach range from $260 million to $440 million, measured as three times the standard deviation of the estimated output. This simulation reveals that the relative costs per record decreases with the number of lost records from $148 per record for smaller breaches analyzed explicitly to $39.5 per record for breaches of 1 million records and further down to $7.0 per record for breaches of 50 million records. Despite this "economy of scale"

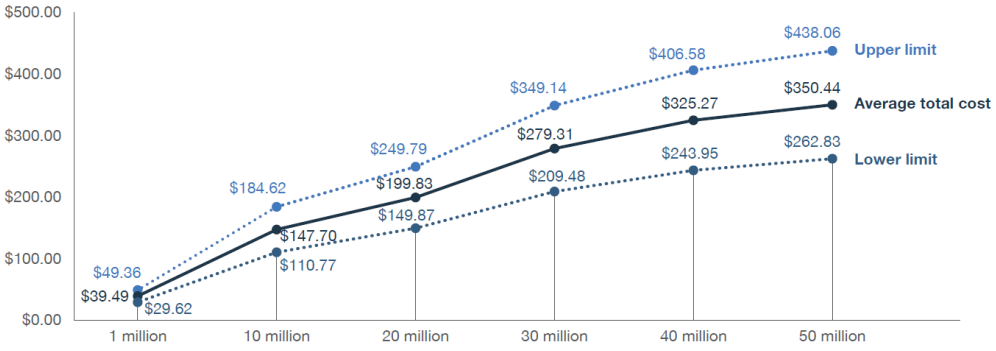

Figure 2: Simulated total cost of mega breaches in $ million. Source: [6].

the absolute costs of mega-breaches can clearly have a dramatic economic impact on the afflicted companies.

Total cost of data breaches at 447 study participants amounts to $1.36 billion (1.4% of the worldwide investment). This is clearly a gross underestimate of the total cost of security incidents since many other types of incidents were not covered in the analysis presented above. Since labor is the main cost factor in the mitigation of security incidents, rich countries like USA, Canada, Europe and Japan are exposed to higher incident costs. Likewise, industries with highly valuable data need to invest more in prevention of security incidents.

1.2 Profits from Security Incidents

Financial profit is one of the main driving forces behind security incidents. In the early history of IT security, incidents were primarily motivated by curiosity or hackers' personal ego. The startling growth of the Internet economy since 1990s has paved the way for integration of security exploitation into a criminal world and gave rise to a complex phenomenon of cybercrime. Modern cybercrime is estimated to generate yearly profits of about $1.5 trillion (see Table 1), which amounts to approximately one third of the world's criminal earnings and more that tenfold of the worldwide investment into cybersecurity.

Modern cybercrime bears a number of substantial features of the capitalist economy. It comprises a large variety of mechanisms for technology development and profit generation and shares typical elements of the legitimate economy: separation of labor, intensive trade, sophisticated payment systems, strong dependence on data, professionalization and skill training, globalization and alternative legal foundations. Cybercrime is characterized by a fast pace of the technological development as well as by highly innovative business processes. Cybercrime is also tightly interconnected with the conventional economy. Companies and nation states not infrequently resort to security exploitation for attaining profits, market advantages as well as political gain.

Crime	Annual Revenues
Illicit and illegal online markets	$860 billion
Trade secret and IP theft	$500 billion
Data theft and trading	$160 billion
Crimeware and cybercrime services	$1.6 billion
Ransomware	$1 billion

Table 1: Annual cybercrime revenue estimates. Source: [12].

1.2.1 Illicit online markets

Illicit online markets are by far the most lucrative source of income for cybercriminals. The paradigm shift brought by e-commerce to the world economy has likewise revolutionized the criminal world. Various features that strongly contributed to the success of e-commerce – e.g., anonymity, cost advantages, convenience and globalization – also proved to be powerful instruments for criminal abuse. Activities of illicit online markets can be classified into the following three categories:

- *Fake e-commerce* refers to various types of abuse that involve misleading users about the nature of the goods or services. Fake web-shops feign the e-commerce functionality by copying the interface and the content of legitimate web-shops. When a user is lured to make a purchase in a fake web-shop, the payment is collected, preferably in an irrevocable way such as a wire transfer, but an order is never fulfilled. Selling of counterfeit goods is another avenue of substantial criminal profits. Since the product authenticity is much more difficult to control in e-commerce, finding and closing fake e-commerce sites has proved to be a big challenge, especially in a cross-border e-commerce ecosystem. Other types of fake e-commerce include selling fake concert or airline tickets, rental contracts, holiday or leisure items, etc. The economy of scale helps fraudsters to reap substantial profits even when individual losses for victims are relatively low and hence their propensity to be deluded is rather high.

- *Abuse of legitimate e-commerce* involves manipulation of various services related to e-commerce. One of the primary targets of this type of abuse are online auctions. With similar fraudulent behavior as in conventional e-commerce – e.g., failure to send goods after the payment is received – online auctions offer attackers additional advantages of high time pressure exerted on users. Another abuse scenario in online auctions is automation of bidding in order to elicit higher prices on items being sold. Such abuse is harder to prevent than in conventional auctions since authenticity of user accounts in online auctions is difficult to control. Online advertisement is another service related to e-commerce that is prone to fraud. The most common type of advertisement fraud is imitation of user clicks which increases the amount of fees to be incurred by an advertiser. Click fraud can be easily implemented by using the special programs on compromised computers which visit websites with certain advertisement and imitate

mouse clicks by executing certain functions in the respective website interface.

- *Online trade with illicit goods* abuses classical anonymity tools by offering anonymous access to illegal services such as drug, firearms and pharmaceuticals markets, electronic piracy, fake documents, mediation of contract killing, etc. The common feature of such illegal marketplaces and services is the use of the "Darknet", i.e., the content intended for anonymous browsing. Such content is addressed by random identifiers with .onion suffix, and tracking down its physical location is highly complex due to the deployment of anonymization mechanisms such as TOR (The Onion Routing). Financial transactions for trading in the Darknet are usually carried out with the help of cryptocurrencies which are well suited for anonymous transfer of funds. Two of the most well known illegal marketplaces in the Darknet were Silk Road and AlphaBay closed by the law enforcement agencies in 2013 and 2017, respectively. The estimated trade volumes on these marketplaces over their life span range from $0.5 billion to $1.2 billion.

Despite a substantial effort to combat illicit online markets, it is unlikely that this problem will abate in the foreseeable future. While technical instruments such as advanced authentication and blockchain can help in detection of fake goods and web-shops, the dual use of strong anonymity mechanisms is likely to remain a major instrument of the cybercrime.

1.2.2 Intellectual Property (IP) and Trade Secret Theft

Gaining profit from the IP theft is a well-known business model in the computer industry. Starting from the early days of personal computers and video recorders, distribution of pirate content on floppy disks, video tapes and compact disks has always been a lucrative business for "small-scale" cybercriminals. The physical distribution of the stolen content, however, has proved to be a dangerous business. Soon after the scale of lost revenues for the computer and film industry had been recognized, stringent copyright protection laws were adopted in the leading western economies. Their effective enforcement has put an end to a trivial monetization of the IP theft.

The emergence of Internet as a ubiquitous communication platform has brought a new dimension to the scale of IP theft. Offering stolen content for download has opened new possibilities for earning profit from such content. While direct monetization via subscriptions still remains rather cumbersome for attackers due to the risk of legal persecution, other potential business models entail advertisement on the pirate content distribution sites as well as planting malware and re-direction of traffic. However, the income of the pirate content distributors and the losses of the creative industry remain hugely asymmetric. The annual advertisement revenues from distribution of the pirate content are estimated at $227 million [2], whereas the losses of the US industry due to the theft of intellectual property are estimated at $320 billion [1].

At the corporate level, the IP theft takes the form of industrial espionage. Numerous incidents have been discovered in the recent decade in which long-lasting infections of corporate networks resulted in a substantial loss of data. Among the prominent companies hit by such attacks are RUAG in Switzerland, ThyssenKrupp in Germany, Daewoo in South Korea, Aramco

in Saudi Arabia as well as numerous universities in the US with close links to the pharmaceutical industry [12]. The profits from the industrial espionage are hard to measure exactly. The estimates put the total worldwide losses from industrial espionage at $445 billion [3], which, accounted for the depreciation of the IP "in transit", results in the projection of profits at $200 billion [12].

1.2.3 Data Theft and Trading

Data leaked as a result of security incidents is a valuable resource for the criminals. It can be abused directly, e.g., via the fraudulent use of stolen credit card data, or offered for sale in the underground economy. Credit card fraud is a well-known problem in the financial industry. The pervasive deployment of credit cards as a payment medium in e-commerce has driven the abuse of credit cards to an unprecedented scale. The problem is further exacerbated by weak authorization mechanisms of credit cards in online deployment. In contrast to purchases at physical locations, authorized by the holder's signature which is difficult to forge, online credit card transactions are authorized only by a 3-digit Card Verification Code (CVC) which is extremely easy to spoof. The criminal profits through the credit card fraud are rather easy to estimate since the financial institutions issuing credit cards are largely liable for the credit card fraud. Yearly losses – and with high probability, revenues of cybercriminals – due to credit card fraud are estimated worldwide at $25 billion [12]. It can be safely assumed that the majority of credit card fraud incidents happen as a result of identity theft in the cyberspace.

Trading of stolen data serves as a stepping-stone for further abuse of such data. Apart from the credit cards, such data includes other financial credentials such as bank account numbers, social security numbers, banking and payment system login credentials, social network login credentials, etc. Assessment of the profits from the credential trade is rather difficult since the actual volume of the respective transactions remains unknown. The best estimate can be obtained by multiplying the number of known credentials offered for sale with their average price. Here, the most valuable source of income are banking and payment system login data with a total value of $116 billion, followed by the credit card data ($15 billion) and login credentials ($0.5 billion) [12]. Other types of data that is commonly traded in the underground economy are diplomas, driver's licences, passports, credit history files and medical records.

1.2.4 Crimeware and Cybercrime Services

Cybercrime could not have reached it current effectiveness and profitability if every actor had to develop his or her own tools. Just like the human civilization has developed trade as a mechanism for exchange of tools needed for productive work, the modern cybercriminal economy is characterized by extensive and effective supply chains for all ingredients of the cybercriminal activity. The main types of commodity involved in such exchange are tools, services and user traffic.

The most popular tools offered in the cybercrimincal black markets are exploits and exploit kits. Their functionality and prices vary largely, ranging from "entry-level" exploit kits

bundling various known exploits ($200-$600) to zero-day exploits for specific products and systems ($5,000-$250,000). Besides exploits, other utilities like DDoS tools or remote access toolkits are available.

In most of the cases, however, an attacker need not even bother about choosing and deploying the specific attack tools. A lot of activities in the cybercriminal world take place as services, with fixed-term leasing rates as well as result-oriented fees. Examples of the services offered for hire are DDoS attacks and botnets, infection of users by placement of web exploits on compromised web sites, exploitation of social media accounts, bullet-proof hosting, etc.

Finally, an important role in the cybercrime economy belongs to traffic re-direction. The bulk of successful exploitation of end-user devices operates via re-directing users with vulnerable web browsers to websites infected with the respective malware. However, popular websites like Facebook, Amazon and Twitter make a substantial effort to protect their users, hence it is rather unlikely that malicious code can be distributed via such sites. As a result, re-direction of users to other, often little known sites with insufficient security is commonly used as a mechanism for initial infection. Traffic re-direction is typically carried out by colluding websites which stealthily re-direct their users to malicious ones. Technically, re-direction can be implemented in a number of ways, e.g., invisible frames, refreshing the web-pages using a special HTTP header, Flash content, etc. Besides infecting end-users, traffic re-direction is also used for other types of profit, e.g., online advertisement fraud and illicit search engine optimization.

1.2.5 Ransomware

The main operational idea of ransomware is to temporarily disrupt operation of the victim system and demand ransom for its repair. Implementation of this idea evolved through several incarnations. Misleading applications, a typical threat for PCs of the mid-2000s, presented themselves as useful utilities, e.g., spyware removal or performance optimization tools, and offered to fix ostensible technical problems for a small fee in the range of $30–$90. The purported problems were largely exaggerated, and in fact, most of misleading applications did not fix anything. As another, almost anectodal monetization technique, some misleading applications requested that victims buy medications in selected online pharmacies and submitted a receipt as a proof of payment, thus earning commission from the pharmacies. A more aggressive variant of misleading applications, fake antivirus, appeared in 2008. These tools pretended to have found a number of security problems on a victim PC and asked users to pay a fee in the range of $40–$100 for removing the infections. Some fake antivirus also offered bogus support services for several years. The financial success of both misleading applications and fake antivirus was low since they failed to disrupt the operation of their victims and were largely ignored or removed by the users.

As a next wave of ransomware, lockers appeared in 2011 with more disruptive functionality. The main technical idea behind lockers is to prevent users from logging into their PCs by tweaking certain registry entries in the Microsoft Windows OS. The stronger impact exerted by lockers on users was reflected in higher ransom, in the range $150–$200, to be paid through electronic cash vouchers. Another strategy to collect ransom fees was to offer "technical support" over expensive premium telephone numbers. Despite the technical superi-

Ransomware	Period	Profit Estimate
Cryptolocker	2013	~$3 million
Cryptowall	2014–16	~$18–$320 million
Locky		$8–$150 million
Cerber		$7 million
WannaCry	2016	$55,000–$140,000
Petya/NotPetya		$10,000

Table 2: Revenues from selected ransomware campaigns. Source: [12].

ority compared to early ransomware, lockers suffered of two major flaws. First, the ransom collection mechanisms proved to be too risky in the face of potential law enforcement. Second, the actions performed by lockers were not irreversible; most of them could be fixed by technically savvy users themselves or were detected by (real) antivirus software.

The turning point in the development of ransomware was understanding of the necessity for disruptions that were hard or impossible to revert without paying a ransom. Since 2013, ransomware switched to the use of industrial strength data encryption as the main disruption instrument. At the same time, the emergence of cryptocurrencies suitable for anonymous payments offered attackers a convenient means for collection of ransom. The key to success of modern cryptolockers is proper key management. Early versions of ransomware made rookie mistakes by using symmetric keys hidden in the malware binary. Such strategies, while effective against end users, were easily botched by the security industry by reverse engineering the malware binary code. More advanced versions of ransomware use standard symmetric encryption algorithms like 3DES, RC4 and AES for bulk encryption while encrypting the session key by strong asymmetric methods such as RSA. As a result, encrypted data can only be restored if the decryption key is known, ergo, after paying a ransom.

The processing of ransom and data recovery are crucial factors for the economic success of ransomware. The challenge of ransom processing techniques is to keep an accurate account of infected devices and payment flows. At the same time, re-use of the same encryption keys for different users must be avoided in order to prevent decryption of data for free. For anonymity reasons, processing of ransom is typically done in the Darknet. This, together with technical difficulties for running the decryption tools, often results in failure of data recovery. Several ransomware campaigns with poorly implemented ransom collection procedures turned into pure destructive instruments inflicting severe economic losses with little financial gain for the attackers. Estimates of the financial success of known ransomware campaigns differ strongly, not the least because the profits were earned in various cryptocurrencies whose exchange rates to fiat currencies has been extremely volatile in the recent years. Estimates earnings for different prominent ransomware campaigns are shown in Table 2. The mediocre profit earned from the WannaCry and Petya/NotPetya campaigns is indicative of the failed data recovery in these ramsomware families. As a result, huge financial losses were inflicted on their victims. The worldwide losses from WannaCry reached $4 billion [8] while those from Petya/NotPetya amounted to $10 billion [9].

1.3 Examples of Security Incidents

1.3.1 Industrial Espionage: RUAG

The security incident discovered in January 2016 at the Swiss defense and aerospace technology company RUAG is one of the few known security incidents classified as industrial espionage. It is characterized by a stealthy compromise of the IT infrastructure, long-term control over the infected systems and a clandestine exfiltration of approximately 23G of data. Investigation of the RUAG incident has been made public by the Swiss Reporting and Analysis Centre for Information Assurance (MELANI) [5].

The RUAG attack is related to a well-known malware familiy known as Epic/Turla/Tavdig. This malware family has been used in several attacks against governmental organizations as well as commercial companies in Europe in the past decade. The timeline of the RUAG incident is illustrated in Figure 3. The incident was discovered on 21.01.2016 when traces of command-and-control activities were found in traffic logs at RUAG's web proxy. The incident was reported to MELANI and the investigation began. During the first week of investigation intensive forensic analysis of logs and disks of the computers suspected to be infected was carried out. As a result of this investigation, malware was discovered and attributed to the Turla family. In February 2016, the investigation was continued by monitoring the suspicious data flows at RUAG. As a result of such monitoring, data exfiltration was observed and further components of the command-and-control infrastructure were discovered. Investigation continued in March-April 2016 until the incident was first mentioned in mass media on 3.5.2016 upon which the attackers' activity was discontinued and no further investigation was possible. An interesting outcome of the investigation was that the first signs of infection were observed by external parties in December 2015. However, these indicators of compromise could not be uniquely attributed to RUAG and hence did not lead to the discovery of the incident. Furthermore, indicators of compromise were also found in logs of several internal servers at RUAG but these were overlooked by its security operations staff. The initial infection presumably took place earlier than September 2014. Its root causes have never been found since no earlier log data was available.

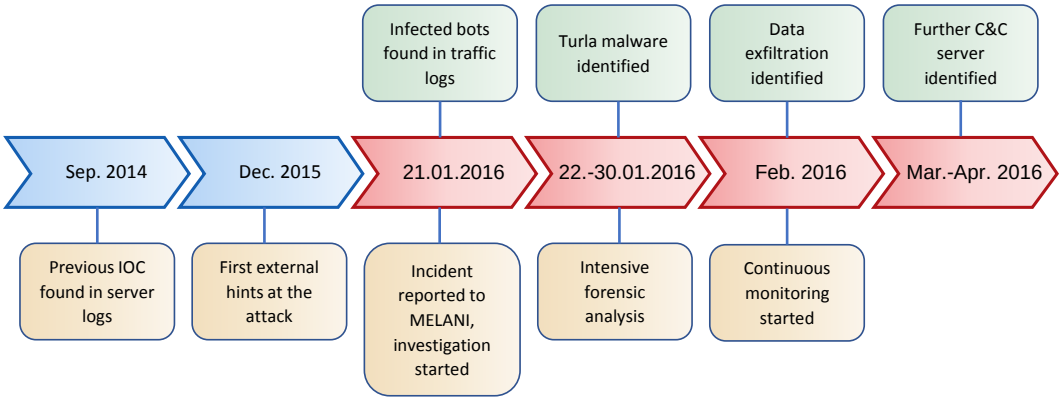

Figure 3: Timeline of the RUAG incident. Source: [5].

Although the precise sequence of events that led to the infection of RUAG is unknown, similar incidents at other victims of the Turla family enable a reconstruction of the general *modus operandi* of the attack. Espionage attacks similar to the RUAG incident typically involve the following stages:

1. *Reconnaisance:* During this phase, the attacker gathers the necessary information about the target, e.g., IP ranges, operation systems and services, as well typical user activities. The ultimate goal of this phase is to set up special websites with malicious content tailored to a targeted user, the so-called *waterholes*.

2. *Infection:* The main challenge of this phase is to find the best exploit for an attack. Based on the information gathered in the reconnaisance phase, an attacker installs suitable exploits at the waterhole and uses social engineering to nudge a targeted user into visiting this website.

3. *Operation:* After an attacker succeeds to infect some user, additional infrastructure needed for achieving the desired information is set up and put in operation. To this end, first, further reconnaisance of internal network services is performed in order to determine targets for lateral movement. Once an attacker understands the technical features of the targeted system, tools for persistent access to the infected system are installed (typically, remote access toolkits, or RATs). In the following steps, attackers escalate their privileges on infected systems and gradually expand their control to the assets of their interest. Finally, the most common operational function of industrial espionage attacks is exfiltration of interesting data.

The specific chain of infection in the RUAG incident is shown in Figure 4. After a user is lured to visit the waterhole website, he or she is re-directed to an infection proxy by means of a JavaScript code camouflaged as a Google Analytics script. The infection proxy checks whether the IP address lies in the range of attacker's interest and if so, delivers a fingerprinting JavaScript code to user's browser. The fingerprinting code, taken from the Browser Exploitation Framework (BEEF), determines whether the user's browser has certain plugins activated and their versions. If a vulnerable plugin is detected, a payload is delived to a user containing an exploit for the corresponding plugin.

Once the exploit is successfully executed on the infected system, a command-and-control (C&C) channel is established with a server under attacker's control and the Turla malware is uploaded and executed on the victim machine. Turla malware contains different components and has undergone several stages in its development, as shown in Figure 5. As a first step after infection, Turla installs and executes a reconnaissance malware, referred to as a *recon tool*, shown as green circles. The main purpose of the recon tool is to reveal further details about the victim's system in order for an attacker to decide if it is worth further infection. If this is the case, the second stage of the Turla malware is installed, referred to as *Carbon* (either a rootkit or a DLL, depending on the version). The Carbon component usually requires administrative rights, hence a privilege escalation attack is executed before it is installed. If Carbon successfully runs on the victim's system, an attacker gains persistent access to it and starts further exploration of and lateral movement in the compromised infrastructure.

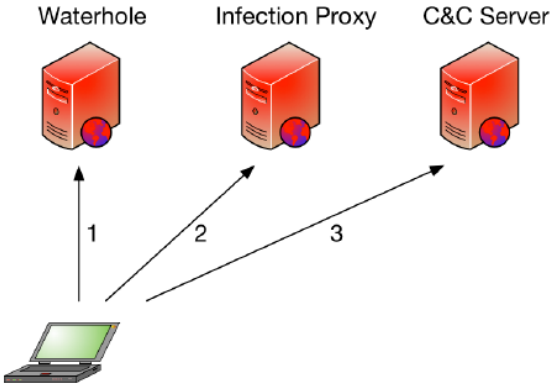

Figure 4: Chain of infection in the RUAG incident. Source: [5].

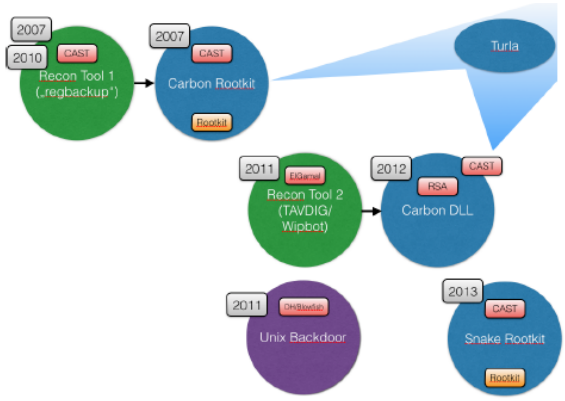

Figure 5: Architecture of the Turla malware family. Source: [5].

To infect further components in the RUAG's infrastructure, several publicly available tools were installed and executed on the infected systems. *Mimikatz*[1] is a popular tool for sniffing access credentials in memory. If it runs during some user authentication dialogue, it can observe and extract various authentication credentials stored temporarily in memory: password hashes, tickets for accessing different resources on the network as well as potentially the Kerberos *Golden Ticket* used for issuing other access tickets. If an attacker succeeds in sniffing the Golden Ticket, it obtains a broad access to other assets in the infected infrastructure. In addition to Mimikatz, attackers also deployed keyloggers and monitoring of cleartext traffic for discovery of login credentials.

A complex and innovative mechanism, shown in Figure 6, was deployed for data exfiltration. After attackers gained access to the machines containing valuable information they were faced by the challenge of transferring their data outside of RUAG since such machines were not directly authorized to establish outbound connections. This problem was solved by

[1] `https://github.com/gentilkiwi/mimikatz`

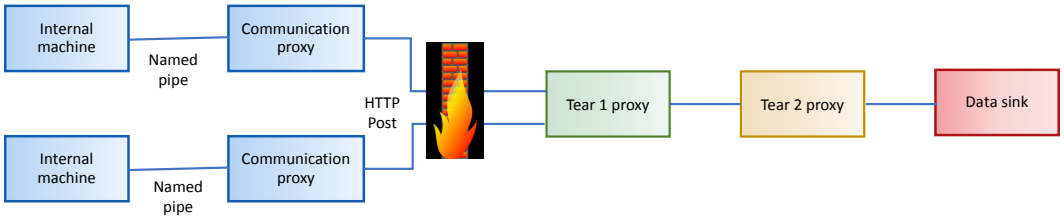

Figure 6: Data exfiltration mechanism in the RUAG incident.

installation of communication proxy tools on the machines with access to the outside world which were able to deliver the data, via several tiers of external proxies to the data sinks under attackers' control. To bypass firewalls at RUAG's Internet gateways, the outbound exfiltration was implemented as POST requests of the HTTP protocol. The internal communication between the data sources and the communication proxies was implemented via *named pipes*, a mechanism for inter-process communication on Windows networks. A named pipe can be created by any process on a Windows machine and encapsulates all necessary primitives required for a TCP/IP communication. By passing the name of a pipe to another process, access to the data can be given to any process on any host, subject to security constraints. By using this mechanism, attackers were able to build an internal peer-to-peer network within RUAG's infrastructure that was used for transferring data from its confidential sources to the communication proxies and forwarding it to the data sinks.

Publication of the investigation outcomes of the RUAG incident sheds valuable insights into technical and operational details of industrial espionage incidents. It reveals that such incidents involved careful planning and patient implementation, combining a large body of sophisticated technical tools. Due to the long timeline of the incident, no evidence could be obtained from the available log data about the exact pattern of the initial infection, which can be only speculatively reconstructed from the analysis of the malware found on the infected machines. The value of the exfiltrated data cannot be estimated since very little data in transit has been observed in the monitoring phase of the investigation. For most of the leaked data, only log entries revealing the time and the amount of the transferred data were available. Due to the complex infrastructure used in the data expfiltration process, it is impossible to track down the perpertrators in the RUAG incident.

1.3.2 Banking Trojan: Carbanak

What would you think if you observe an ATM machine dispensing cash by itself, with nobody standing in front if it to withdraw cash? You might think that perhaps the machine is infected by malware. Indeed, most of the ATM machines are built on legacy Windows systems (often Windows XP or older) and can be infected via USB ports after some physical tampering, which has been used in several security incidents, e.g., [13]. In spring 2014, several such misbehaving ATM machines were spotted in Kiev, which led their operators to request an investigation by malware experts at Kaspersky Labs. The upshot of their analysis was: no malware on ATM machines. Instructions to dispense cash were sent over legitimate VPN channels from internal ATM networks. Instead of ATM machines the entire operational infrastruc-

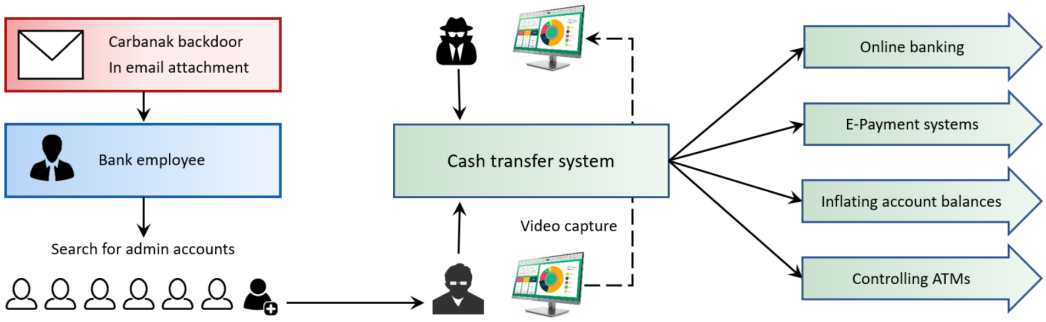

Figure 7: Conceptual overview of the Carbanak trojan.

ture of the respective banks was compromised and various kinds of illegal transactions were executed by criminals. The ensuing investigation, the results of which were published by Kaspersky Labs [4], revealed a sophisticated technical and operational attack inflicting an unprecedented damage in the class of financial malware. The total financial losses of "The Great Bank Robbery", which affected more than 100 banks around the globe, are believed to be in excess of $1 billion. How did the attackers manage to infiltrate and abuse extremely sensitive banking systems?

The conceptual overview of main technical and operational aspects of the Carbanak trojan deployed in these attacks is presented in Figure 7. The starting point of Carbanak attacks were spear phishing emails with Microsoft Word 97-–2003 (.doc) or Control Panel (.cpl) files attached. The doc files exploited well known vulnerabilities in Microsoft Office (CVE-2012-0158 and CVE-2013-3906) or Microsoft Word (CVE-2014-1761). The text of the messages appeared legitimate for bank employees and was often sent from compromised accounts of their colleagues. The text of one of Carbanak's spearphishing emails translated from Russian (the attack originally targeted Russian banks) is shown in Figure 8. When a user opened an attachement with an unpatched Word or Office application, the exploit successfully ran on a victim system and installed the payload of the Carbanak trojan. To establish a persistent control over the infected system, the Carbanak trojan implemented the command-and-control (C&C) functionality using a custom protocol incapsulated in HTTP requests. It also enabled opening Remote Desktop (RDP) connections to the infected system.

Carbanak used a variety of tools for its lateral movement. Its preferred means of controlling the infected system was the Ammyy Admin remote administration tool, probably because it

```
Good Day!
I send you our contact details
 The amount of deposit 32 million rubles and 00 kopecks, for a period of 366
days,% year---end contribution term
 Sincerely, Sergey Kuznetsov;
+ 7 (953) 3413178
f205f @ mail.ru
```

Figure 8: Example of a phishing email used in Carbanak. Source: [4].

was whitelisted on target systems. Traces of other tools such as Mimikatz, PsExec and Metasploit were also found by Kaspersky researchers on infected systems. The goal of the lateral movement was to obtain login credentials for administrators of cash transfer systems. Once attackers were able to login as administrators, they had to understand the operation of such systems and learn the necessary controls in order to execute financial transactions.

In addition to traditional tools such as keyloggers deployed in trojan malware, Carbanak used "video recordings" in order for attackers to learn the operation of cash transfer systems. The recording system took screen shots at regular time intervals and transferred them in a highly compressed form to the C&C servers. Using the intelligence gained from these monitoring techniques, attackers developed an operational picture of the victim's workflow, tools and practices. Such knowledge enabled them to earn profits from executing various kinds of fraudulent transactions:

- *Online banking.* For manual control of online banking, transactions can be executed from back-end cash transfer systems. Using such transactions, attackers could directly transfer money to accounts under their control.
- *Electronic payment systems.* Such systems are used for transferring funds between banks. Payments were made in such systems to attackers' banks "in the name of" legitimate payment system operators.
- *Account balance inflation.* By inserting fake transactions directly into the back-end databases, attackers could change balances for the accounts they opened in victim banks and cash-out the gains.
- *ATM control.* Attackers issued commands to dispense cash from certain ATM machines at certain times which was expected to be picked up by their accomplices.

Similar to industrial espionage attacks, the Carbanak trojan has exhibited an amazing versatility in infection and lateral movement, which was instrumental for attackers' ability to maintain control over the victim systems for a prolonged time. Carbanak's remarkable feature is that it exploited no vulnerabilities in cash transfer systems or ATM machines. All that was needed to inflict humongous financial losses was to obtain access credentials for the crucial control points in such systems and to patiently learn how to use them.

1.3.3 Ransomware: Samsam

Unlike the majority of ransomware threats, the SamSam ransomware, named after a first file discovered by security researchers, used a very primitive distribution mechanism while being very effective in spreading inside a compromised network and encrypting valuable files. SamSam's initial penetration strategy used brute-forcing of weak passwords to log into Windows machines over the Remote Desktop Protocol. Alternatively, earlier versions of SamSam used a JBOSS application server vulnerability to move to their new targets. After logging in, the attacker would use privilege escalation to identify login credentials for further assets on the network such as servers, endpoints or other targets. After the attacker controlled enough

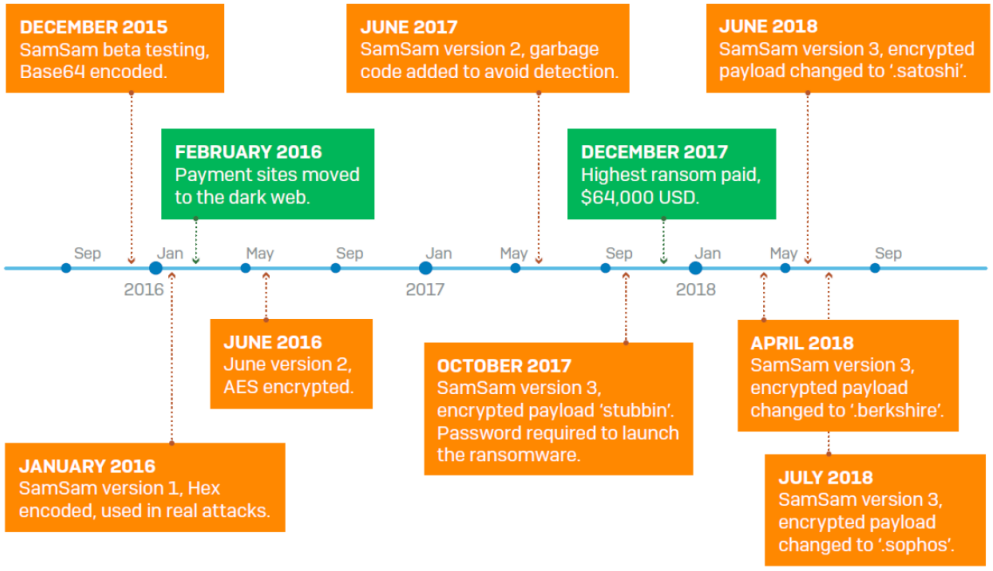

Figure 9: Evolution of Samsam ransomware. Source: [7].

assets, an encryption process would start (usually outside of working hours) and first corrupt a prioritized list of files before encrypting the remaining data. A particular feature of SamSam was the lack of automation tools: the entire attack was performed manually by an attacker who carefully controlled the victim's machine.

During the course of its evolution, presented in Figure 9, SamSam steadily improved its reliability and stealthiness. In the early versions of SamSam the ransom note was simply encoded using Bas64 or hex encodings. Subsequent versions released in June 2016 deployed AES for encrypting the ransom note in the file to make analysis more difficult. Starting from June 2017, the attacker started to add garbage code to make the analysis of captured samples more difficult. In the last version of SamSam, the payload was split across two files, one of which decrypts the payload of the other one using a password manually entered by the attacker. The decryption password was changed at regular time intervals.

A complex procedure deployed for collection of ransom enabled an attacker to control the flow of infections and payments. For each victim, a separate payment site was created on a Darknet[2]. The address of the payment site, the Bitcoin payment address as well as the ransom amount were displayed to a victim in a ransom note. The full ransom usually amounted to 7 BTC and had to be paid within a week from the infection. If the user did not pay within the deadline the payment site was taken offline but could be re-opened for an additional fee of 0.5 BTC. The victim had other payment options. For 0.8 BTC it was possible to receive the decryption key for only the initial target compromised by Samsam (excluding the victims of its lateral movement). For 3.5 BTC a user could decrypt half of the infected machines. After the payment was made the victim had to inform the attacker by visiting the victim-specific payment site on the Darknet and communicating with an attacker by means of a

[2]Earlier versions hosted the payments sites on anonyme.com or anonymous WordPress sites.

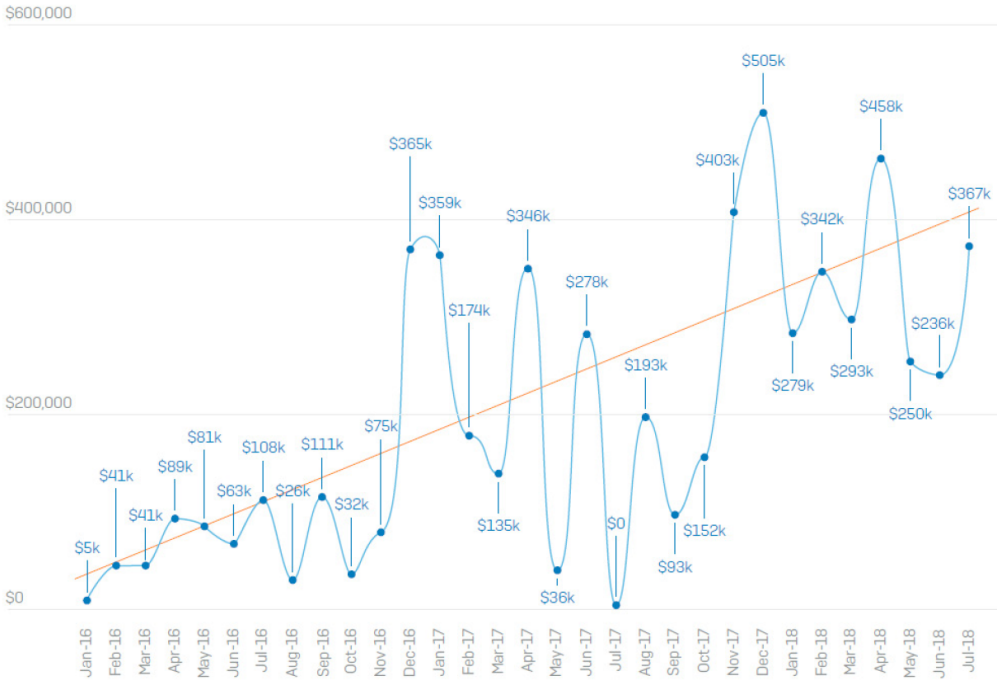

Figure 10: SamSam payment history. Source: [7].

simple HTTP form. After the payment had been received, an attacker would upload a ZIP file with a decryption key to the payment site that could be downloaded by the victim. The file would contain two routines for decrypting and deleting the encrypted files, a private key needed for decryption as well as a text file containing the instructions for running decryption. It is the victim's responsibility to run the tools provided by the attacker to restore the data.

For anonymity, SamSam demands ransom in Bitcoins. Despite the anonymity of Bitcoin transactions the flow of money can be tracked since all transactions are recorded in a public ledger. As a result, even though the identity of the ransom senders and recipients cannot be immediately established, a thorough analysis of the transaction chain gives valuable insights into the operational side of the attack. It further enables a rather accurate estimate of SamSam's operational profit. In the course of analysis, 157 unique Bitcoin addresses have been identified that received ransom payments as well as further 89 Bitcoin addresses that were mentioned in ransom notes but never received payments. According to Neutrino, a firm that specializes on blockchain analysis, all these addresses can be associated with three wallets, each controlled by a single person. The transaction flow shown in Figure 10 exhibits a clear linear grown over the time. The US$ equivalents of the observed Bitcoin transactions were computed using the Bitcoin values at the time of conversion to US$ (which can be tracked as deletion of the respective Bitcoin units) or at the conversion rate corresponding to the end of study [7]. The total profit of the observed operations of SamSam is estimated at $5.9 million.

Given the specific features of SamSam, in particular, its mostly manual modus operandi, one may wonder what this reveals about the identity of attacker(s)? While very little can be

learned about SamSam's authors in underground fori, there is evidence to believe that the entire development and operation was carried out by a single person. It can be seen in the text of ransom messages that the attacker was not a native English speaker. Such texts contain consistent grammar and spelling mistakes as well as currency formatting like 15,000$ not typical for English-speaking countries. Another consistent linguistic feature of both ransom notes and payment instructions is capitalization of letters in words following commas. This suggests that an attack may originate from a country in whose language commas can be easily confused with periods. Analysis of time stamps for compilation of malware files does not deliver any clues to attacker's origin since they lack time zone information. However, it could be observed that most of the files were compiled during regular "working hours" between 9am and 1am.

Chapter 2

Security Goals

The notion of information security is defined in terms of *security goals*. Each security goal describes a broad class of properties characterizing essential features that are required in order for a system to be considered secure. At the top level, the following security goals are defined:

- *Confidentiality:* information is not disclosed to unauthorized recipients.
- *Integrity:* information is not modified in an unauthorized way.
- *Availability:* information is accessible to all authorized users.

Some additional crucial properties of information systems are often pointed out in relation to the top-level security goals:

- *Privacy:* individuals control which information is collected about their activity and how this information is used. Privacy is related to confidentiality in that it prevents unauthorized use of information about user activity.
- *Authenticity:* the origin of information is defined and cannot be manipulated. Authenticity is understood as the integrity of the origin of information.
- *Accountability:* the usage of information is traced and cannot be manipulated. Accountability implies the existence and the integrity of the monitoring of the of information systems' usage.

The above mentioned security goals and additional properties are discussed in more detail in the following sections.

2.1 Confidentiality

Prevention of an unauthorized disclosure of information has been known as a security goal for millennia. Ancient Greeks as well as Romans used various devices and methods to scramble physical messages carried by their couriers. The need to keep communication secret – as well as the opponent's desire to break such secrecy – has motivated the development of *cryptographic tools*, the main instrument for enforcing confidentiality. Confidentiality needs to be enforced any time information is transmitted over an insecure channel. A classical example

of such communication is radio transmission, as well as any type of wireless communication that can be intercepted by others. Wired or optical networks are also often considered to be insecure unless communication parties not ensure that they share a dedicated physical channel that cannot be eavesdropped.

Besides unauthorized reading of data, disclosure of any further information about communication, e.g., who communicates with whom or at what time, may also be undesirable. Internet users may be interested in keeping their browsing history hidden from other parties, and vice versa, the Internet industry has a strong interest in understanding users' behavior for better marketing or other commercial insights gained from user data. Preventing unauthorized leakage of information about users is the essential subject of privacy. It is enforced by a variety of instruments using cryptographic techniques to either completely destroy information about user indentities (anonymization) or to temporarily remove user identities from data (pseudonymization).

2.2 Integrity

Data tampering may occur in a variety of use-cases in information systems. Modified bank account data in a payment transaction will cause money to be transferred to a wrong account. A falsified amount of a incoming transaction will increase an account balance in a fraudulent way. Changing the content of program code downloaded from the Internet may infect a computer with malicious code. Tweaking the URL of a bank website may lead to the leakage of users' access credentials for sensitive financial services.

The mechanisms for enforcing the integrity of information rest on the computation and verification of message authentication codes. Such code is sensitive to any modification of the original data. Message authentication code is computed before and after the transmission of a message. If a discrepancy is found between the two codes an integrity violation is reported. The security of this mechanism crucially depends on the inability of an attacker to either find a modified message with the same authentication code or to substitute the original message authentication code with the one corresponding to the tampered message. These requirements can be met by a special type of transformations known as *cryptographic hash functions*.

Verification of integrity plays a crucial role in authentication mechanisms, i.e., determination of the identity of an instance involved in the exchange of information. Authentication is a prerequisite for deciding whether a given instance is allowed to access information. If an attacker succeeds in spoofing the identity of an authorized user, he or she may gain unauthorized access to information. In the networking context, authentication plays a crucial role in ensuring the identity of communication end-points. A attacker may claim an identity of a certain end-point and forward the user traffic to that end-point while observing the network traffic. In such attack, often referred to as *man-in-the-middle*, an attacker may even succeed to observe encrypted content by decrypting traffic (which is possible if an attacker negotiates a common encryption key with a user) and re-encrypting the traffic forwarded to the genuine end-point. To ensure the integrity of user identities the classical authentication techniques based on shared secrets, e.g., passwords, are enhanced with cryptographic tools, e.g., digital

certificates, or two-factor authentication based on the possession of additional artifacts, e.g., biometric features, special apps on a smartphone or cryptographic token generators.

Another important use-case for integrity verification is enforcement of accountability. To provide irrefutable evidence for crucial events in information systems the collected event trace must be protected against manipulation. In an exemplary scenario, a user may place an order in an online store but later rescind the payment after receiving the goods by claiming that he or she never ordered anything. To protect a vendor in this case, the record of orders has to be maintained in such a way that the user identity is verified in an irrefutable way and the record itself cannot be manipulated by mischievous users. These tasks can be addressed by standard techniques for integrity verification as well as their modern derivatives, e.g., blockchain.

2.3 Availability

From the operational perspective, availability is the most crucial security requirement. Operational failure of information systems may lead to dramatic financial losses as well as endanger the physical security of technical infrastructures and human life. Needless to say that the high value associated with availability of information systems motivates miscreants to abuse it in order to gain monetary benefits. The most common abuse scenario is to execute or to threaten a denial-of-service attack and to demand ransom for the recovery from failure or for not staging the attack. Technical realizations of denial-of-service attacks are quite diverse. At the network level, such attacks can be implemented by flooding servers with connection requests or data in order to deplete their computational resources or network bandwidth. Another type of denial-of-service attacks involves sending of malformed data causing processing errors and leading to server crashes. At the system level, the most prevalent form of denial-of-service attacks is ransomware discussed in detail in Section 1.2.5.

A broad arsenal of technical solutions is needed to assure availability of information systems. Adequate design with respect to scalability of computing and networking resources is the first essential step in assuring availability. Contingency planning, such as re-configuration of networking infrastructures, dynamic outsourcing into the cloud and backup solutions, are other important provisions for availability. Further technical instruments enable detection and recovery of network-level denial-of-service attacks. Due to the simple operational patterns of such attacks rules can be designed for close-to-real-time detection of their onset. Novel denial-of-service attacks can also be detected by observing abnormal features of network traffic characteristics. Once a denial-of-service attack is discovered and its sources and targets are identified, appropriate countermeasures can be initiated, such as traffic cleaning and dynamic resource allocation.

Chapter 3

Security Management Principles

Given the economic considerations related to information security presented in Chapter 1 it is natural to approach the task of security management as a risk optimization problem. By doing so one can consider the whole spectrum of potential measures in a single decision-making framework based on assessment of security risks. In this chapter, we elucidate the main constituents of security risks and outline typical processes of security management.

3.1 Security Risk

From the operational perspective, security can be defined as the state in which the risk is below the certain acceptable value. The residual security risk can be carried by an organization or an individual themselves, or it can be transferred to an insurance policy against security losses. It is essential for both scenarios to develop a procedure for estimation of security risk.

Mathematically, risk is defined as a sum of costs weighted by their probabilities, for all possible events:

$$R = \sum_{x} C(x)P(x). \tag{3.1}$$

To compute the risk, the cost C and the probability P of incurring this cost must be determined for every event x.

The cost component of the risk in the security context is related to organization's *assets*. The latter comprise physical devices (end-hosts, servers, communication lines, etc.), software, data, intellectual property and know-how, business reputation, customer base, etc. Identification and documentation of assets is a starting point for estimation of security risk. Asset valuation can be carried out according to the replacement value or the lost business value. While the former method is straightforward, computation of the lost business value can be highly non-trivial. An example of such computation was presented in Section 1.1 for data breach incidents. Similar computations based on prior experience in the organization or within a specific industry can be used as a guideline for asset valuation. Another rule of thumb that can be helpful in asset valuation is to consider potential disruptions to organization's operation due to the loss of a specific asset. Such disruptions can be measured in the number of productivity units, e.g., person-days, lost due to a potential security incident, and the cost can be estimated by adding up the respective productivity values.

The probability component of the security risk is an interplay of vulnerabilities and threats. *Vulnerability* is any weakness of a system that can be exploited to damage assets. Technical

vulnerabilities arise as a result of programming errors: lack of boundary checks during memory access, lack of input validation, logic errors, memory management errors, any many other types of errors. Configuration vulnerabilities stem from inappropriate configuration settings, e.g., listening to unnecessary ports or using weak cryptographic methods. Administration vulnerabilities include weak settings related to system administration, such as default passwords, excessive access privileges or flaws in user authentication. Manual identification of vulnerabilities is largely infeasible; however, various tools exist for automatic identification of known vulnerabilities, e.g., nmap, nessus and Webinspect. Funding novel vulnerabilities (often referred to as "zero-day" vulnerabilities) is currently beyond the reach of automation tools; discovery of such vulnerabilities can have a large impact on the entire computer industry, as it happened with the "Spectre" and the "Meltdown" bugs.

To enable the assessment of security risks, vulnerabilities are documented and stored in vulnerability databases. Examples of vulnerability databases are BugTraq, CVE and NVD. Severity of vulnerabilities is rated on a numeric scale from 1 to 10, or in three categorical classes, high, low and medium. Numerical rating according to the Common Vulnerability Scoring System (CVSS) introduced in 2005 is used in the CVE database. CVSS uses a standard computation formula that combines six factors characterizing the impact of a vulnerability: access vector, attack complexity, necessity of authentication, impact to confidentiality, integrity and availability. According to the statistics collected for the CVE database, about 13% of vulnerabilities have the CVSS score of 9 to 10, whereas the remaining scores are almost equally distributed in the range of 4 to 8. Categorical ratings of vulnerabilities include critical (automatic exploitation possible), medium (exploitation can be prevented by configuration) and low (exploitation is difficult and the benefit is limited).

Vulnerabilities as such do not contribute to the security risk. To measure their impact on security, it is necessary to address *threats*, i.e, actions carried by attackers to exploit vulnerabilities. Such actions depend on vulnerabilities but may also depend on external conditions, e.g., possession of valid login credentials or certain access rights. Threat preconditions may not be satisfied at the beginning of an attack but may be met as a result of intermediate attack steps. To capture the dependence of different threats on each other, *attack trees* can be constructed that graphically represent all possible attack strategies and the interdependency of particular attack steps. An example of an attack tree is shown in Figure 11.

Security risk can rarely be exactly calculated using the formal definition presented in Equation 3.1. The main problem lies in computing the probability of a successful attack. The intermediate steps of attacks are highly dependent, therefore probabilities cannot be computed as products of probabilities of individual steps. As an additional complication, a successful attack against some security mechanism may lead to a complete violation of assumptions made in the design of other security mechanisms. An obvious example is password-based authentication which assumes that passwords are unknown to third parties. If, however, an attacker is able read a password from computer memory by exploiting a vulnerability in authentication software, he or she can gain full access to a service that requires password-based authentication.

In practice, security risk estimation is carried out *qualitatively* by a joint assessment of the likelihood and the impact of different threats by means of a single *risk matrix*. In a typical

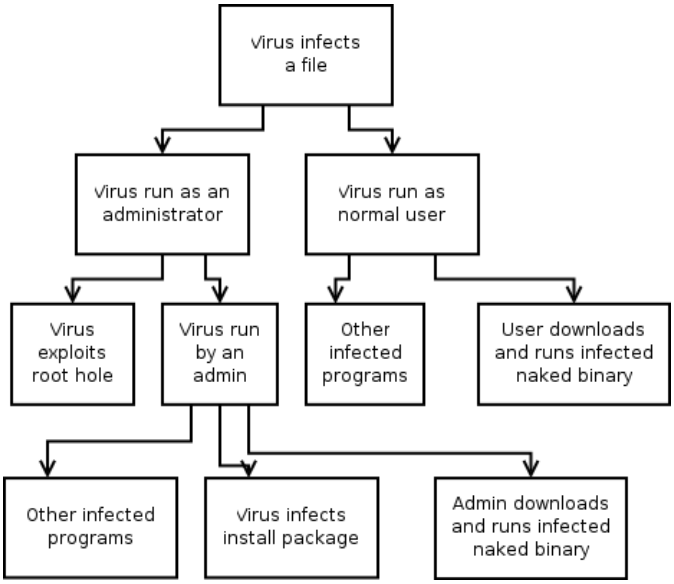

Figure 11: An example of an attack tree.

risk matrix shown in Figure 12, ordinal values[1] are defined for the impact and the occurrence of threats. These values constitute the horizontal and the vertical axis of the matrix and are mapped to the specific risk categories, e.g., low, medium, high and very high. The main advantage of the risk matrix is that the impact and the occurrence are easy to estimate for various threats based on the empirical evidence. Unambiguous mapping of risk categories to impact/occurrence values enables one to assign certain actions to each risk category and thus to prioritize the implementation of appropriate countermeasures.

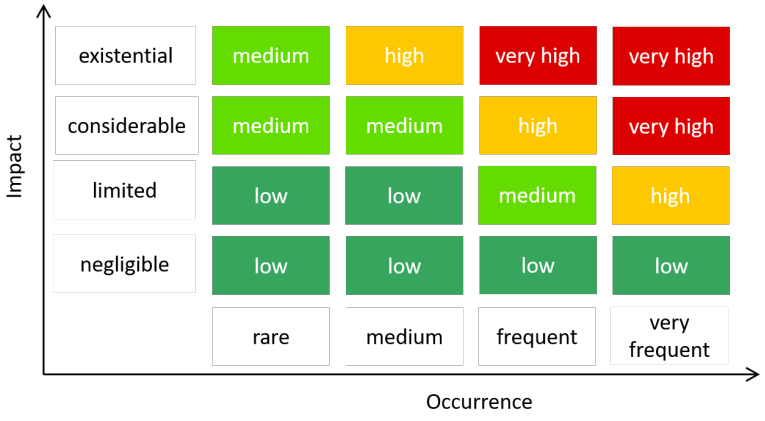

Figure 12: An example of a risk matrix.

[1] *Ordinal* variables are categorical variables with the sense of order between different categories.

3.2 Security Management Workflow

Security management defines the process for achieving and maintaining key security goals: confidentiality, integrity and availability. This process involves the definition of the security strategy, identification of security risks, selection and implementation of security controls[2], monitoring of the operation, detection and response to incidents and re-assessment of security risks. The overall workflow of a typical security management process is presented in Figure 13.

The starting point of the security management is *system characterization*. At this step the key assets of the organization's information system are identified and documented. The system characterization involves the definition of the system's mission, its functional requirements, interfaces, as well as the description of users and management procedures. As a result of this step the system's boundaries are defined, its functions are elucidated, and critical modules and data are identified.

During the *threat identification* stage the potential threats to the system are analyzed. A crucial task of threat identification is understanding of threat sources. Based on the system operation history, the potential threat actors are identified and, if possible, the motivation for their activity is documented. Last but not least, external threat intelligence sources which report global security threats in various industries can be studied, and the applicable threats can be analyzed within the specific organization's context.

Vulnerability identification is focused on specific weaknesses of a given information system. Various sources can be used for vulnerability identification. Technical vulnerabilities can be identified for various assets by studying vendor security alerts as well as industry recommendations. Automatic vulnerability scanners can also be helpful in discovering novel

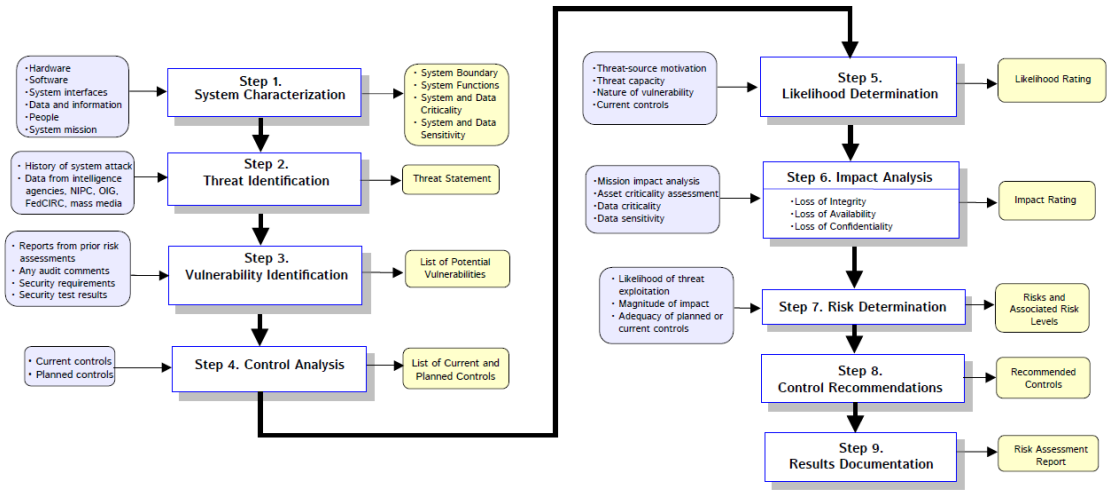

Figure 13: Security management workflow.

[2]The notion of *control* refers to any technical instrument or organizational measure that enables to decrease certain security risks.

technical vulnerabilities. For identification of more complex technical vulnerabilities, penetration testing, i.e., practical exploitation by dedicated security experts, is getting increasingly common and in some industries is even mandated by regulation. Non-technical vulnerabilities can be identified by specifying security requirements and comparing them with the implemented operational procedures. Common practice is also to correlate the identified threats and vulnerabilities in a single list, which enables one to detect the crucial weaknesses of a system requiring a prioritized action.

Analysis of controls delivers further input for security risk assessment. Controls constitute the key instrument of security management. They enable mitigation of the risk of an incident even if a vulnerability cannot be removed. For example, if a server is vulnerable to privilege escalation enabling unauthorized access, this vulnerability can be mitigated by restricting the access to this server to only selected computers with sufficient access control mechanisms. Controls can also be used to improve security monitoring. By defining and implementing the policies for external access to the system, monitoring accuracy can be improved by ensuring that all traffic passes through the controlled entry points into the system. To enable a systematic analysis and implementation of controls, security standards define compendia of controls with detailed descriptions of the threats they can mitigate.

The analysis carried out in steps 2 – 4 enables *likelihood determination* for each threat. The key factors affecting the likelihood of successful attacks are the criticality of a vulnerability, the degree of attacker's motivation (and potentially, monetary incentives), the existence and the effectiveness of mitigating controls. As mentioned in Section 3.1, the likelihood is usually defined in terms of ordinal values, e.g., low, medium, high, very high.

The next major step in the security management workflow is *impact analysis*. This step crucially depends on the system mission and functionality. The impact is determined based on the criticality of the system for business operation as well as the sensitivity of data. In many cases, the impact can be measured in terms of the absolute loss values, as shown by the study presented in Section 1.1. For other kinds of impact that cannot be measured directly, ordinal values are assigned based on the empirical consideration.

The purpose of *risk determination* is to assess the level of risk to information systems. As explained in Section 3.1, this step entails assigning the risk values to specific threat/vulnerability pairs based on their likelihood and impact. The risk assignment may take into account existing controls, attacker motivation as well as human factors in the organization.

Control recommendations constitute the ultimate step of security management. Based on the estimated risk values suitable controls are selected in order to attain an acceptable value of the residual risk. Selection of controls is affected by a number of factors: technical feasibility and effectiveness, legislative and regulatory constrains, organizational aspects of their implementation, operational impact. Last but not least, cost-benefit analysis plays an important role in the selection of controls. Higher priority is assigned to those controls that attain better reduction of security risk at lower implementation costs.

Documentation of intermediate findings and final results is a self-evident requirement of security management. Documentation provides a foundation for the implementation of agreed measures, but it also delivers input for subsequent iterations of the security management cycle. The risk assessment report serves as the basis for management decisions and resource allocation by organization's senior management.

3.3 Security Strategy

The holistic view of all issues relevant for security management is digested in the form of *security strategy*. The main purpose of the security strategy is to provide a reference for all decisions related to security management. Security strategy comprises the following elements:

- *Security policy* describes the desired behavior of an information system. It defines key security requirements, documents all security-related policies and procedures, describes the roles and responsibilities of the relevant stakeholders, documents the existing controls and their operation. Development of the security policy should take into consideration the desired level of security, ease of use, impact of security measures on productivity, implementation costs, risk of system failure as well as legal and regulatory requirements. Security policy constitutes a business decision and must be approved by organization's senior management.

- *Security implementation* comprises procedures and tools for prevention and detection of security incidents as well as for response to and recovery from security incidents. Preventive mechanisms ensure that specified security requirements are effectively implemented. Detection mechanisms enable the discovery of attack symptoms and – should the preventive mechanisms fail – early detection of indicators of compromise. Response mechanisms are necessary for stopping ongoing attacks to prevent further damage. Recovery mechanisms are deployed for restoring potentially lost data and resuming the normal operation of a system.

- *Security assessment* measures the success of security management. As a basic instrument of assessment, *assurance* techniques are deployed to verify the implementation of security requirements. During the system design stage, assurance is used to guarantee that the system design matches security requirements. At deployment, assurance verifies the correct implementation of design elements. Assurance cannot deliver a formal proof of security; however, it can increase the confidence in a proper implementation of security mechanisms. As a stronger proof of security – albeit still of empirical nature – security *evaluation* is carried out as part of security assessment. Evaluation involves examining products and systems in terms of measurable criteria. For example, the accuracy of detection mechanisms can be evaluated by running them on data with known ground truth. Effectiveness of personnel training can be measured by performing theoretical examinations or practical tests. Overall effectiveness of security mechanisms can be evaluated by exposing the system to penetration testing.

Bibliography

[1] The Report of the Commission on the Theft of the American Intellectural Property, 2013.

[2] Good money gone bad: Digital thieves and the hijacking of the online ad business, 2014.

[3] Net losses: Estimating the global cost of cybercrime, 2014.

[4] Carbanak APT: The Great Bank Robbery. Technical report, Kaspersky Labs, 2015.

[5] APT Case TUAG. Technical report, MELANI, 2016.

[6] 2018 Cost of a Data Breach Study: Global Overview, 2018.

[7] SamSam: The (Almost) Six Million Dollar Ransomware. Technical report, Sophos, 2018.

[8] Charles Cooper. WannaCry: Lessons Learned 1 Year Later, 2018.

[9] Andy Greenberg. The Untold Story of NotPetya, the Most Devastating Cyberattack in History, 2018.

[10] Ponemon Institute. Staffing the IT security function in the age of automation: A study of organizations in the United States, United Kingdom and APAC, 2019.

[11] Avivah Litan, Elizabeth Kim, John A. Wheeler, Christian Canales, Dale Gardner, Deborah Kish, Ruggero Contu, and Sid Deshpande. Forecast: Information security, worldwide, 2016-2022, 2Q18 update, 2018.

[12] Michael McGuire. Into The Web of Profit: An in-depth study of cybercrime, criminals and money, 2018.

[13] Chris Vallance. Cash machines raided with infected usb sticks, 2013.