Information Classification in Context

Matej Kačic I Senior IT Security Consultant
Maroš Barabas I Head of Security Technologies Division
Hana Vystavělová I Head of Compliance and Risk Services Division

Abstrakt: Článek předkládá praktické poznatky z implementace klasifikace informací a její návaznosti na bezpečnostní normy, požadavky zákona o kybernetické bezpečnosti a regulativy GDPR. Zaměříme se na přínosy klasifikace informací v rámci procesu budování bezpečnosti v podnikovém prostředí a následnou implementaci bezpečnostních technologií na detekci a prevenci úniku informací a bezpečnostního monitoringu. Uvedeme příklady správného nasazení klasifikace a ukážeme, jak vhodný nástroj zvyšuje bezpečnostní podvědomí (security awarness) a vynutitelnost procesu klasifikace v prostředí společnosti.

Klíčová slova: GDPR, klasifikace informací, bezpečnost

Abstract: This article presents real findings based on the implementation of information classification, its connection to safety standards, and requirements set by the Cybersecurity Act and the GDPR regulations. We shall focus on the benefits the classification of information brings, if applied during the process of building security in the enterprise environment and on the subsequent implementation of security technologies used for detection and prevention of information leaks and security monitoring. Number of examples will be listed where correct implementation of classification was applied and we will show you how a suitable tool increases security awareness and enforceability of the classification process in the company environment.

Keywords: GDPR, information classification, security

Introduction

Already in the ancient times, people endeavored to protect their information from falling into the hands of unauthorized persons. This is why many encryption mechanisms were invented, all with a singular goal - to protect sensitive information from one's enemy. Even though the information that were subject to classification were previously mostly of military character, today, in times of information and communication technologies development, no company can exist without some means used to protect its know-how in a form of information. And as the volume of processed information increases, along with the expenses spent on protection against the more and more sophisticated attacks, the need to sort, i.e. to classify, the information increases as well.

Classification models were first used for government and military purposes (e.g. Bell – LaPadula model [8] or Biba model [9], which aimed also at protection of integrity). These categorized the information from the most sensitive levels (Top Secret), all the way down to the least classified (Public or Unclassified). Access to the classified information was then given to the users on basis of passing an appropriate security clearance.

The ISO/IEC 27001 standard (or its predecessor BS 7799-2, respectively) ensured massive expansion of classification systems also outside the government and military sectors. This standard regards the classification of assets as one of the basic pillars of information security. The standard requires all processed information to be classified within the selected level

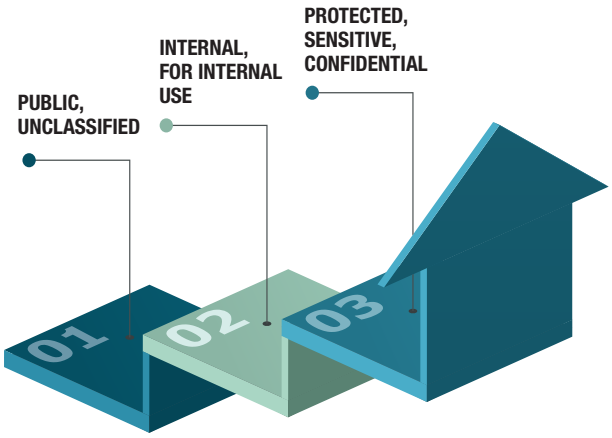
(with regard to its value, legal requirements, sensitivity, and criticality). Further, a procedure for marking and handling the information has to be set, as well as rules for storage and manipulation with the data.

Also, the Czech legislation has the classification of information in mind. In the Act No. 412/2005 Coll., on the Protection of Classified Information and on the Security Authority, there are defined four categories of information - Top Secret, Secret, Confidential, and Restricted. The law also clearly states the requirements for protection of such information and the conditions under which it is possible to access them. The Cybersecurity Act, or the follow-up cybersecurity regulation (No. 316/2014 Coll.), respectively, is yet another example. The regulation states that the affected subjects must also set a security policy for classification of assets. It is common to meet the categorization pursuant to Act No. 101/2000 Coll., on the Protection of Personal Data. Apart from defining the class of personal data, it defines so-called sensitive personal data as well. Stricter measures have to be taken when collecting, handling, and protecting such data. The above-mentioned Act has been currently subject of intense discussions, because the General Data Protection Regulation (further only GDPR) shall come into force on 25th May 2018. This regulation is regarded as a revolution in the personal data protection. GDPR causes great anguish among the companies processing personal data, because when breaching some of the selected provisions of this European regulation, the penalty may reach up to EUR 20 million. And if the transgressor is a company, it may amount up to 4% of its total turnover worldwide. This definitely sounds like a solid argument for finding out what type of personal data the company processes, how it is handled, and where exactly it may be found in the information system. To achieve this goal, apart from other methods, a classification system can be implemented, or possibly other suitable security technologies, as we shall demonstrate in the following lines.

What the Classification Schemes Look Like?

According to our experience, it is not difficult to define a classification scheme. Problem starts when the scheme has to be carried through into the day-to-day company life. Using an internal directive, the organization defines 3 to 5 levels/categories, into which it classifies the information. Confidentiality and access to such information is usually used as a classification criterion. The following names and characteristics are commonly used when naming the categories:

Public, Unclassified - indicates the lowest level, which is commonly described as information intended for publishing (company website, marketing and



information materials, leaflets, public source information). If unauthorized users access this category, there is no negative impact on the organization;

Internal, For Internal Use – described as information accessible to all employees or to a selected group, but not accessible to anyone outside the organization without approval given by the authorized person, who is usually the information owner. Typical examples of this category are internal information important for company operation, internal organization and company activities, internal directives, etc.;

Protected, Sensitive, Confidential – information that require higher level of protection, intended for selected individuals or groups of employees only (management, sales department, system administrators etc.), these must not be freely accessible either to other employees, nor to subjects outside of the company. Personal data, strategic information, and business information all fall under this category.

For schemes where there are more than 3 levels, an additional separate category is usually defined for personal data or there are more levels with higher protection degree (Trade Secret, Strictly Confidential, Highly Sensitive, etc.).

Marking rules are also part of the definition – i.e. it is defined which categories have to be visibly labeled, and how. In real life, a label is inserted into the header/footer of the document, or the body or subject of the e-mail, into the information system output, etc.

Implementation Pitfalls

We encounter quite a number of issues when implementing the schemes into practice. On one side, there is a low security awareness – the users do not know how to classify and label, or their classification is incorrect. In extreme cases, the employees are not able to even name all the defined levels, or they label the documents with their own (sometimes even made-up)

levels. Regular trainings are the solution. But these may be time-consuming and sometimes quite burdensome from an organizational point of view.

When it comes to more complex (multi-level) schemes, it may be difficult for the users to see the difference between the individual levels: Is this particular document Sensitive or Highly Sensitive? Where to classify the document holding both strategic business information and personal data at the same time? Therefore, a great emphasis has to be put on a very precise definition of each category.

The reluctance to use the classification system, namely to label the individual documents with the relevant category may complicate the implementation process as well. It is usually caused by unclear or unenforced responsibility for document labeling, or simply by the employees' and their supervisors' negligence. Inserting the labels directly into organization templates usually helps. But a lot of data is generated outside these formal templates (informal documents, e-mails, system data, scanned data, etc.)

A large number of historic, and therefore unclassified, data is usually also seen as a complication. No reciprocal manual classification is executed if that is the case, but simply a date is defined, from which onward all newly generated information must be classified. Documents that originated before this date are then classified when they are opened or further reused.

Based on our experience with implementing the classification schemes into practice, we recommend to consider the following best practices:

- Define simple and unambiguous classification scheme. Subsequently, set the rules for inserting labels and protection within the whole life-cycle for each category (when and how to encrypt, where to store, how to send it outside the protected internal perimeter, how to dispose of it). Remember to include not only digital, but also physical form of information.
- Regular training for the users; use model examples during the trainings.
- See to thorough control and enforcement of observation of the scheme (management inspection, checking during internal audits, random checks executed by the Security Manager). Take appropriate corrective or disciplinary actions when internal rules are breached.
- Incorporate the labels into all company templates, as well as to all internal systems and applications in use (if possible).
- Define the responsibilities for classification, including the name of the person, whom the users can reach in case they have any issues when using the classification scheme (Security Manager, Security Administrator, etc.).
- Set the classification for the most often used information groups clearly, in order to make the decision for users easy; for example - project documents – Protected, work contracts – Personal Data, company website – Public, SAP – Sensitive, etc.

The following chapter shall deal with classification scheme implementation and enforcement thereof.

How to Enforce Classification of Information?

We recommend to use a suitable tool for information classification, to enforce the best practices within the organization. When implementing such a tool, there is no need for continuous training of users, for checking if the systems is adhered to, and bonus, efficiency of tools such as DLP (Data Loss Prevention) is increased. Integration with other systems (apart from the above mentioned DLP, it may be for example RMS and SIEM) adds to overall awareness in the organization about all the places where sensitive information is stored, as well as about manipulation with such data. Such knowledge may be a very valuable input when adhering to the measures required by the GDPR. Yet another undisputed benefit is the possibility to automatically classify historically created documents en masse.

The suitable tool should be configurable to a high degree, in order to meet maximum of the requirements arising from the different security policies, company directives, standards, laws, or regulations. Main requirements for such tool include:

- Simple definition of own classification categories.
- Setting and changing the look of the visual label – the label must be set according to the company graphic templates. As is often the case, the user must be able to choose, which graphic template should be used for the particular document.
- Setting the look of the DLP label in the document metadata. This string needs to be defined independently of the language used and, at the same time, it must be unique within the company.
- Setting enforced encryption based on the classification category - different options of encryption must be supported, for example native MS Office encryption, zip, or interconnection with access management systems.
- Recording audit logs about document classification.
- Granular settings based on a membership in a particular domain group.
- Possibility to (pre)classify document templates.
- Logging the classification process events directly into the systems such se SIEM in order to draw correlations that identify the users who are not using the classification, location and work with protected data, or if the document has been re-classified from a higher classification category to a lower category.

Classifying Data of „in use" Type

„In use" type of data is the kind of data, which is used actively. Among the examples of such data may be mentioned newly created documents, current documents in active use, or data stored in e-mails or calendars. These types of data can be classified by using a simple tool, which will classify the document at the moment when it is opened, saved, or printed out. Classification means insertion of a visual label together with insertion of an appropriate string into the document metadata.

The tool should enable the user to re-classify on purpose (i.e. change of the classification level), while the previous classification level is recorded in the document history, together with the name of the user executing the change. The user has to list the reasons for change in classification during re-classification.

When the tool is in operation, we recognize several classification modes:

- Passive – user can access the classification interface in the menu, but he/she is not notified in any way about the obligation to use classification when saving a document / sending an e-mail.
- Optional – in case the document is not classified when it is being saved / printed out / an e-mail is being sent-out, then a window is activated that offers the user the classification menu and notifies him/her about the requirement.
- Mandatory – in case the document is not classified when it is being saved / printed out/ an e-mail is being sent-out, then a window is activated that offers the user the classification menu and notifies him/her about the obligation to classify it. It is not possible to save the document / to send the e-mail without proper classification.

The whole process of data classification should function in such way as to ensure that the tool checks if the document being saved or printed out has been already classified, or not. In case it has not been classified yet, it will offer the user an opportunity to classify it, based on the selected classification mode.

Classification of e-mails is done in a similar way, it is mandatory to classify only outgoing e-mails. Incoming e-mails are classified automatically on basis of the previously classified e-mail conversation, or manually during opening or saving the e-mail.

Classifying Historic Data („data in-rest")

As we already mentioned in the theoretical introduction, historic and unclassified data is a great complication. There may exist millions of such documents that we rate among the „data in rest" in one company only.

The only possible solution to classify historic data correctly it to define a set of rules, which shall identify the relevant classification category of the documents based on their contents. These rules are defined on basis of an expert knowledge of an analyst knowledgeable about the company situation to great detail. For example, there is a rule set that is counting a number of different birth certificate numbers listed in one document. And if

the number of these birth certificate numbers is bigger than the set limit – e.g. 5, then we are dealing with a list of persons, or customers, respectively.

If we want to generalize the example above, then we cannot do without searching in the document, which can be based for example on regular expressions, together with specification of number of occurrences. Then we need to assign significance to each result of the search (searched for string and number of occurrences). Further, we have to combine them into larger and more complex rules, using propositional logic. That is how we acquire quite a strong means for classification of historic data.

It is quite common that we are not able to design the rules to be totally unambiguous, which means that we need to have the tool that tells the system to propose the classification and, at the same time, to hand the proposal over to the data owner for execution of classification based on his/her decision.

In the end, there is nothing else left than to classify the given data and save it in a way that will not change the date of last modification.

Classification in Middleware and in Cloud

When implementing classification scheme in the enterprise environment, we often encounter the situation when the documents are not created directly on the user's workstation, but are created in an information system or in cloud. Example of this situation may be the case when the user fills-in an on-line form and the information system subsequently generates a document for him/her, or when an external supplier uploads scanned documents into the data warehouse. Document classification is very difficult in cloud environment as well. The main representatives of cloud solutions are Microsoft Office 365 or Google Docs.

To solve this situation, we can use interconnection of information systems, middleware, or cloud services with the classification service, which will classify the document generated or created in cloud, add the metadata, and generate a SIEM event. Integration to the classification service is provided by a web REST api.

Connecting Classification to Technologies

The basic requirement for implementation of information classification is to successfully identify (or find) the information, define the classification rules, as well as adaptation of processes for manipulation with information, together with monitoring and protection thereof. There are several technologies available that help organizations to implement classification and to enforce the processes defined.

Manual analysis of infrastructure and company processes may be used during the process of identifying the areas (such as data storages) where the information is stored. Other possibility to identify such areas is to use a semi-automated method, while using some of the available technologies. DLP technologies and classification tools are able to scan through data storages and a network data stream, where they are automatically searching within the contents of documents and are able to identify the searched-for information according to the defined rules. The process of identification of these rules may be quite complicated and it is often the case that the correct categorization may not be provided by technologies (i.e. by regular expressions). Therefore, it is necessary to often check the outputs manually. After they are found and potentially checked, it is possible to classify and label them automatically.

DLP technologies also help to enforce the set rules for manipulation with data and to protect sensitive information. Correct definition of DLP policies is the crucial step that must be taken to prevent an intentional, as well as unintentional attack, but a complex analysis must precede it. Continuous classification of data (files, documents, e-mails, etc.), executed especially by the end-users, contributes to clear identification of sensitive information by detecting classification labels or tags, not only on the workstations, but also in storages and within the network data stream.

Systems not prepared for classification and labeling of data may present a problem when protecting the sensitive information. These are the systems such as databases, various monitoring tools, or portable devices. It is possible to protect the data stored in database systems for example with FW database technology, which can check access to protected data on basis of defined policies. Nevertheless, the presentation layer over the database (e.g. on-line application) must be protected separately, as well as the database access, data transfer (communication), etc. Protecting data on the users' end-stations and on portable devices is very complicated, especially when it comes to devices that are leaving the secure company infrastructure. Of course, it is crucial to keep to the best practices applied in the workstation protection (AV, DLP client, strong security policy, and user awareness). RMS systems may help as well, since they can manage access rights to this information based on the classification and the set policies.

It may be more fitting to monitor any unauthorized access than to set preventive access restrictions. In case the data is classified, labeling the data and defining correct DLP policies brings the possibility to monitor any activities connected to the sensitive data (e.g. creation of documents, sending sensitive information via e-mail, copying, access to shared storages, etc.). It is possible to

obtain this information from the DLP system, the classification tool, the RMS system, etc. Based on the defined rules, security incidents resulting in leakage and misuse of protected data can be identified and evaluated, especially in case of implementation of the SIEM (Security Incident and Event Management) tool.

Conclusion

When implementing classification of information in the enterprise environment, we always aim to cover all places where the documents are being created – be it a workstation, cloud, or documents generated automatically from the information system. We always make an effort to enforce classification at all these places, which means that users have no other choice than to classify the data in question; while historic data has a specific status in the classification process. In any case, it is not possible to enforce or automate classification of historic data without a proper tool that should act as sort of a connector between the classification of information, data loss, and evaluation of incidents in the SIEM system.

Acknowledgment

This article was written as part of the IT Reliability and Security project (FIT-S-14-2486). This work was created with support of the Ministry of Education, Youth and Sports, within the National Sustainability Programme II (NPU II) and the project IT4Innovations excellence in science - LQ1602.

References:

[1] BS 7799-2:1999: Information Security Management -- Part 2: Specification for Information Security Management Systems, (April 2001).

[2] ISO/IEC 27001:2013: Information technology — Security techniques — Information security management systems — Requirements, International Organization for Standardization, Geneva, Switzerland.

[3] Act No. 412/2005 Coll., on the Protection of Classified Information and on the Security Authority.

[4] Act No. 181/2014 Coll., on Cybersecurity and on the Amendments of the Related Acts.

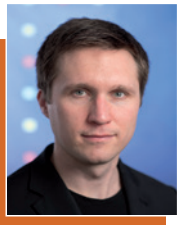
[5] Regulation No. 316/2014 Coll. on Security Measures, Cyber Security Incidents and Reactive Measures, and on Determination of the Requirements for the Applications in the Field of Cyber Security.

[6] Act No. 101/2000 Coll., on the Protection of Classified Information and on the Amendments of some Acts.

[7] Regulation (EU) 2016/679 General Data Protection Regulation – General regulation on protection of personal data.

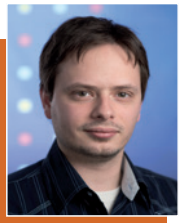
[8] Bell, D. E. - LaPadula, L. J.: Secure Computer Systems: Unified Exposition and Multics Interpretation, Report MTR-2997 Rev. 1, MITRE Corporation, Bedford, Mass, 1976.

[9] Biba, K. J. „Integrity Considerations for Secure Computer Systems", MTR-3153, The Mitre Corporation, June 1975.



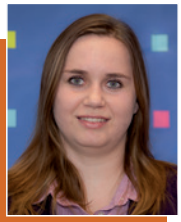
Matej Kačic

Absolvent FIT VUT v Brne, súčasnosti študent doktorského štúdia tej istej fakulty, kde sa zaoberá sa bezpečnosťou informačných systémov a bezdrôtových sietí. Je členom skupiny Security@FIT na VUT v Brne, aktívne prednáša na akademickej pôde a venuje sa publikačnej činnosti. Od roku 2013 pracuje ako Security Consultant v spoločnosti AEC v divízii technológií, kde zastáva hlavne rolu architekta bezpečnosti. Medzi jeho hlavné oblasti patrí návrh a audit bezpečnej sieťovej infraštruktúry. Medzi jeho nosné technológie patria NG firewally, NBA a sandboxy. V posledných dvoch rokoch zastáva rolu product ownera nástroja DocTag určeného pre klasifikáciu dokumentov.



Ing. Maroš Barabas

Vyštudoval doktorský študijný program na Fakulte informačních technologií, VUT v Brně so špecializáciou na bezpečnosť a zameraním na behaviorálne aspekty sieťovej komunikácie, kde doteraz pôsobí ako výskumný pracovník. V roku 2006 nastúpil do novo vznikajúcej českej pobočky spoločnosti Red Hat, kde od roku 2008 pôsobil v tíme zameranom na bezpečnosť Linuxových systémov a to až do roku 2011. Od začiatku roku 2012 pracoval na pozícii IT Security Consultant v spoločnosti AEC a od roku 2014 vedie tím špecialistov zameraných na bezpečnostné technológie. Maroš sa profesionálne venuje hlavne oblasti bezpečnostných technológií so zameraním na detekciu malware, sieťovú analýzu, bezpečnosť informačných systémov a sietí, štandardy pre automatizáciu bezpečnosti, bezpečnostné nástroje a technológie, sieťovú behaviorálnu analýzu, návrh, implementáciu a audit bezpečnostných architektúr, systémov a technológií. V neposlednom rade sa zapája do šírenia bezpečnostného povedomia 17 odbornými a vedeckými publikáciami na bezpečnostných konferenciách a v odborných a vedeckých časopisoch.



Mgr. Hana Vystavělová

Absolventka Fakulty informatiky Masarykovy univerzity v Brně. Od roku 2005 se věnuje informační bezpečnosti ve společnosti AEC na pozici konzultanta informační bezpečnosti. Zaměřuje se především na problematiku řízení informační bezpečnosti, analýzy a řízení rizik, dokumentace, bezpečnosti dle standardů řady ISO/IEC 27000 a souvisejících oblasti. Od nástupu do společnosti se podílí na řadě projektů pro zákazníky z oblasti veřejné správy, telekomunikací i finančního sektoru.