Securing the Flow
Data Flow Analysis with Operational Node Structures

Michael Meinig and Christoph Meinel
Chair of Internet Technologies and Systems,
Hasso-Plattner-Institute (HPI), University of Potsdam, 14482 Potsdam, Germany

Keywords: Risk Management, Security Awareness, Threat Awareness, Threat Modelling, Vulnerability Analysis.

Abstract: After land, sea, air and space, cyberspace has become the fifth domain of warfare. Organizations recognize the need for protecting confidential, secret - classified – information. Competitors and adversaries turn to illegal methods to obtain classified information. They try to gain a competitive advantage or close a technological gap as well as reduce dependencies on others. Classified information involves facts, subject matters or knowledge needing to be kept secret, regardless of the way in which the information is depicted. In networks with different security classifications a direct physical connection is not allowed. Consequently the possibility of coupling different security domains in affected organizations must be checked comprehensively under security aspects. In this paper we present a new security approach that helps to identify threats at transitions and security zones on valid data flow paths. It can be used to display security challenges within organizations using classified information such as governmental or military organizations. The methodology also incorporates new attributes for data flows in connected systems or processes.

1 INTRODUCTION

Organizations spend money on security solutions to protect their classified information. They implement policies derived from regulations and laws to prevent the loss of confidentiality, availability and integrity of their information. Still, due to operational requirements exceptions to policies may be allowed and authorized (Eckstein, 2015).

An employee who is often away on business trips will possibly be allowed to use his USB port to download presentations and at the same time upload information he has collected on his journeys with his laptop into the business network. The authorization of these exceptions is often purely based on the business demand for the individual user. This poses the question of how a permission for an exception does influence the security risk for an organization.

New communication forms like instant messaging, Voice over IP or blogs and storage possibilities such as cloud computing are used in organizations besides to those traditional ones like email, telephone, USB flash drive or hard drives (Gordon, 2007). Attacks like spear-fishing and social engineering are popular attack methods which worm themselves deliberately into networks (Trendlabs, 2012). Even the savviest IT professionals are sometimes not aware of the difference between a real and fake event (Mah, 2017). How can we consider these new conditions at an early stage of the development process?

Over the last few years many organizations across all different sectors have confirmed data breaches (Cyberedge, 2015). This data loss or leakage caused through theft or loss by internal offenders and trusted third parties can be intentional or unintentional (Infowatch, 2016). The costs incurred of data breaches amounted in millions for organizations (Ponemon, 2016). The modelling of those threats is more cost-saving if applied as early as possible in development process. The costs for changes to diagrams are significantly lower than changes to a system in production (Torr, 2005).

1.1 Problem Statement

There are organizations, i.e. in the NATO or in the EU or in a nation, which have different operational node structures and which are dealing with classified data or having different security classifications zones.

Meinig, M. and Meinel, C.
Securing the Flow - Data Flow Analysis with Operational Node Structures.
DOI: 10.5220/0006570302410250
In *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018)*, pages 241-250
ISBN: 978-989-758-282-0

The problem under investigation may be informally stated as follows: given a set of "security classes" corresponding to classes of information, and a specification of possible paths by which information can flow among them, construct a mechanism which clusters all objects of one security class into one security classification zone and then reduces and identifies all connections to only those connections which flow between these security classification zones. A direct physical connection between these security classification zones is not allowed but due to security breaches the loss of confidentiality, availability and integrity is immanent and hence it follows that valid paths are used to perform illegal and undesirable operations. If classified information is published unauthorized it has an impact on many stakeholders because it could pose a threat, a damage or disadvantage to the interests of the stakeholder. Given the described situation, it is important to have certifications and modelling approaches that identify such issues.

1.2 Organization and Results of This Paper

The paper starts with background information and a review of related work in which we discuss how our solution advances the state-of-the-art. In the next chapter we define a data flow diagram with a special security focus. Thereupon we use the security data flow diagram to conduct an analysis which identifies operational node structures that are affected most by the threat of losing confidentiality, availability and integrity of their classified information while exchanging data between networks with different security levels. Finally in the chapter conclusion and future work we state the conclusions reached by this research and propose areas for future study.

The important contributions of this paper are summarized below: A model of data flows and security is developed. It is used for identifying data flows between different security classification zones. An analysis of these data flows and the security problem itself leads to the conclusion that some operational nodes connected to those data flows are endangered more to the loss of confidentiality, availability and integrity than others which makes it possible to prioritize them for mitigation efforts. The model is an enhancement of previous work on the security problem. It enables to make statements about the probability that illegal and undesirable operations have been executed on valid data flow paths.

2 RELATED WORK

2.1 Information and Data Flow Models

There has been much work on information flow and information flow models. D.E. Bell and J. LaPadula introduced a security model for computer systems which protects the confidentiality of information using a system of enforced rules. Information of a higher protection level can neither be read nor transferred to a lower protection level (Bell, 1976). Kenneth J. Biba came up with a security model which addresses the integrity of data, checking read and write access in a computer system (Biba, 1976). D.E. Denning studied mechanisms that assure secure information flow in a computer system (Denning, 1976). She proposed a lattice model for secure information flow. Its structure evolves from different security classes and is validated by the semantics of an information flow. Based on this model, D.E. Denning and P.J. Denning demonstrated a certification mechanism for statically verifying the secure information flow in a program (Denning, 1977). A.C. Myers and B. Liskov introduced a model for controlling information flow in systems with mutual distrust and decentralized authority. In this model it is possible to share information with distrusted code and furthermore to decide whom the information is shared to (Myers, 1997). J. Rushby suggested that secure systems should be designed as distributed systems in which security is achieved partly through the physical separation of their individual components and partly through the mediation of trusted functions performed within some of those components (Rushby, 1981). W.S. Harrison, et al. presented a joint research effort between academia, industry, and government called Multiple Independent Levels of Security and Safety (MILS) in order to develop and implement a high-assurance, real-time architecture for embedded systems. The goal of the MILS architecture is to ensure that all system security policies are non-bypassable, evaluatable, always invoked, and tamper-proof (Harrison, 2005).

All these papers and approaches provide a basis with methods and models for secure systems and their information flows. Secure information flows ensure that sensitive information is not leaked to unauthorized entities during program execution. (Bell, 1976), (Biba, 1976), (Denning, 1976), (Myers, 1997), (Rushby, 1981), (Harrison, 2005) do not focus on the problem that the implementation or use of a secure system may -- due to implementation errors or various attack vectors, such as social

engineering -- be not secure at all. Consequently there are security threats or vulnerabilities which can be used to carry out illegal and unwished operations over valid paths across network borders.

2.2 Information and Data Flow Analysis

Data flow diagrams or similar techniques for representing the flow within systems, such as flowcharts have been present in literature since the seventies (Gane, 1977), (DeMarco, 1978), (Yourdon, 1989).

Microsoft uses data flow diagrams within the Microsoft's threat modeling methodology which is part of the Security Development Lifecycle (SDL). A. Shostack described a decade of experience of threat modeling at Microsoft. He notices that DFDs are very data-centric and the analysis is focused on "the right thing" (Shostack, 2008). K. Schmidt, et al. present a security analysis approach that helps to identify and prioritize security issues in automotive architecture. This approach uses data flow diagrams for a structured threat analysis and risk assessment in a security-oriented development process (Schmidt et al., 2014). K. Schmidt, et al. use communication zones where entities are able to communicate directly with each other, due to a shared communication layer.

In our presented approach the model of data flows is connected with security attributes. Instead of communication zones security classification zones are utilized. Due to the focused problem and the existing security restrictions security classification zones are physically separated and a direct communication is not possible. Our analysis of these data flows and the security problem itself leads to the conclusion that some operational nodes connected to those data flows are endangered more to the loss of confidentiality, availability and integrity than others which makes it possible to prioritize them for mitigation efforts.

3 DEFINITION DFDsec

Due to the dominance of the Unified Modelling Language (UML) in software engineering and the Systems Modelling Language (SysML) in systems engineering, data flow diagrams are not widely used in the field. UML specification defines two major kinds of UML diagram: structure and behavior diagram. Structure diagrams present the static structure of the system and its parts on different abstraction and implementation levels and how they are related to each other. Structure diagrams are not using time related concepts, do not show the details of dynamic behavior. For modelling the behavior of a system the UML standard uses activity diagrams. Activity diagrams are neither adequately capable of representing data flows. Instead they focus on the transitions between sequencing activities.

A data flow diagram contains processes, data stores, external entities and data flows. In this approach we add security classifications as security attributes to the data flows. Furthermore we use these security classifications, which are applied by governments and military, to create zones around the original data flow diagram elements. This is similar to the work by Microsoft and Schmidt et al. discussed before. The boundaries of security classification zones are necessary because a direct physical connection is for security reasons not allowed. Within or between security classification zones security principles and protective measures are applied which shall prevent classified information from the threat of loss of confidentiality, availability and integrity.

In a military or governmental environment, people, documents and information can receive two types of formal security designations: one is the classification or clearance (unclassified, restricted, confidential, secret, and top secret are usual) and the other one is a formal category (such as Nuclear, NATO, EU, DEU and Crypto). Such a pair we call a "security level". We define the security level as "I(c)", where I is the formal category and (c) is the classification or clearance. For Germany an example would be "I(c) = DEU (RESTRICTED)". Instead of using specific security classifications increasing numerical numbers are used for c which are elements of integers Z, such as:

- PUBLIC $= -1$
- UNCLASSIFIED $= 0$
- RESTRICTED $= 1$
- CONFIDENTIAL $= 2$
- SECRET $= 3$
- TOP SECRET $= 4$

In a NATO or EU context these classifications would be altered accordingly. The following four cases, within for example the sector of government and enforcement agencies, of information flows in a security data flow diagram (DFDsec) consider the possible resulting situations, which are:

- data is sent within the same security classification zone (i.e. confidential

information - technical know-how) - allowed flow of information
- data is sent from a lower to a higher security classification zone (i.e. personal identifiable or health information) - forbidden backflow of information
- data is sent from a higher to lower security classification zone (i.e. departure or destination records) - released backflow of information
- data is not sent from a higher classification to a public network (i.e. state secrets) - forbidden flow of information

A security data flow diagram is defined by

$$DFDsec = <P_{(n)}, E_{(n)}, S_{(n)}, F_{I(z)}, I(z)> \quad (1)$$

where:

- Processes $P_{(n)}$ from 1..n represent normally in a standard data flow diagram a task in the system that processes data or performs some action based on data. In our model we link them with entities which are performing activities by operating on incoming data and potentially producing output (e.g. operational nodes generating confidential information). The reason for this linkage is simply to identify the originator of the activity. This is necessary for a later risk evaluation. Processes are represented by circular shapes in the figures.
- External Entities $E_{(n)}$ from 1..n are interactors outside the inspected system which upon the system depends. They can be the source or destination of information. They can be part of another or even a whole security classification zone. External Entities $E_{(n)}$ cannot be used when functional requirements are defined because at this stage the future system is not defined. Hence it is not possible to define which of the Processes $P_{(n)}$ will be handled within or outside a system. Rectangular shapes represent external entities in the figures.
- Data Stores $S_{(n)}$ from 1..n are physical or logical repository for storing or retrieving data (e.g. users, data bases, file system). Data Stores $S_{(n)}$ cannot be used when functional requirements are defined because at this stage the future system is not defined. Hence it is not yet clear which physical or logical repository will be used for the future project. Open-ended, rectangular shapes represent data stores in the figures.
- Data Flows $F_{I(z)}$ are defined by

$$F_{I(z)} = <A, FB, RB, 0> \quad (2)$$

There are information flows between processes, external entities, data stores and security classification zones. According to the situation in which they are used, they may either be allowed flows (A), forbidden backflows (FB), released flows (RB) and forbidden flows (0), specific attributes which are added to the data flows. Data flows are represented by arrows pointing in the direction of the flow and are highlighted. Data flows $F_{I(z)}$ which flow within one and the same security classification zone I(x) are allowed if $z \leq x$.
- Security Classification Zones I(z) are the specific security level which incorporates processes, data stores, data flows, external entities belonging to it, where I is a set of formal categories (e.g. Nuclear, NATO, EU, DEU and Crypto) and (z) is a set of classifications or clearances (e.g. unclassified, restricted, confidential, secret, and top secret). To distinguish between two security classification zones we use I(x) and I(y), x and y ϵ Z. Security classification zones are represented by circles surrounding all elements of one security classification zone in the figures.

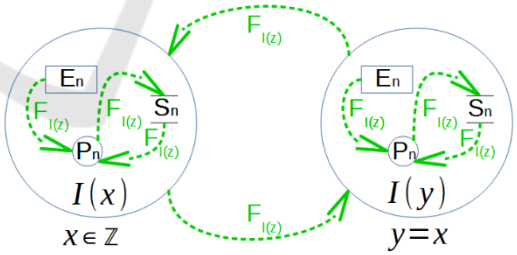

Figure 1: Data Flow "Allowed".

Figure 1 represents the first situation where data is sent within the same security classification zone. An information exchange is possible within an area of public or unclassified information as well as where the principle "Need-to-know" is applied[1]. There are no other restrictions within the same

[1]It is applied only within the same security classification and only when the security classification is higher than RESTRICTED.

security area. These are allowed data flows $F_{I(z)}$ are flows from zone I(x) to zone I(y), while x ϵ Z, y=x and z = classification.

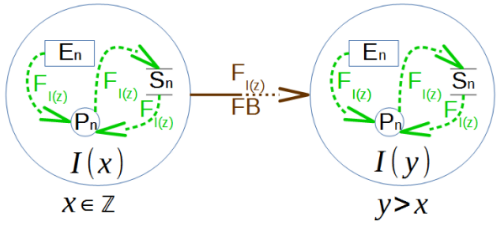

Figure 2: Data flow "Forbidden Backflow (FB)".

Figure 2 depicts the second situation where data flows are only allowed in one direction. An information backflow is forbidden in this case. Allowed data flows $F_{I(z)}$ in one direction are flows from zone I(x), where x ϵ Z and z = classification and z ≤ x, to a higher security classification zone I(y), where y > x. A high security gateway such as a data diode realizes such a connection (Genua, 2016). They are tagged $F_{I(z)}$.

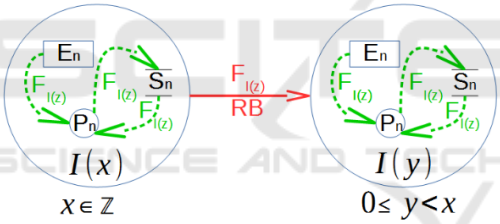

Figure 3: Data flow "Released Backflow (RB)".

Figure 3 illustrates the third situation where data exchange between different security classifications is only possible if the information backflow is released. The data flow runs from a higher security classification to a lower one but not lower then UNCLASSIFIED. Allowed data flows $F_{I(z)}$ with a released backflow are flows from zone I(x) to zone I(y), where x ϵ Z, 0 ≤ y < x and z = classification and z ≤ y. A high security gateway such as a red-black gateway which allows precise content monitoring and controlling data flows between networks with different security classifications implements such a connection (Infodas, 2016).

Figure 4 demonstrates the fourth situation of a data exchange. But in this case the data flow is forbidden due to legal reasons. It is prohibited to send classified or unclassified information to a public network. Forbidden data flows $F_{I(z)}$ are flows from zone I(x) to zone I(y), where x ϵ Z, x > y, y = − 1 (−1 represents the public network) and z = classification and z > y.

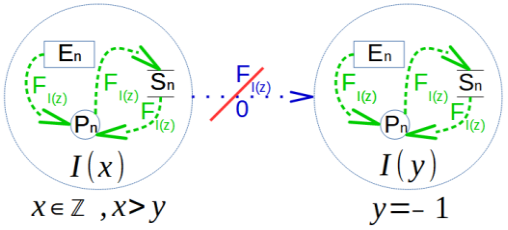

Figure 4: Data flow is strictly forbidden.

4 ANALYSIS OF A DFDsec

In this section we use the DFDsec to conduct an analysis which identifies operational node structures that are affected most by the threat of losing confidentiality, availability and integrity of their classified information while exchanging data between networks with different security levels. In this analysis we combine the three fundamental aims of IT security (BSI, 2008): confidentiality, availability and integrity with the possible security levels.

There are various reports and surveys about attack types (i.e. Malware, Web based attacks, Denial of service, Physical manipulation/ damage/ theft/ loss or Phishing), attack vectors (i.e. Cyber-criminals, Insiders, Nation States, Corporations, Hacktivists, Cyber-fighters, Cyber-terrorists Script kiddies), sectors (i.e. banking and finance, government and enforcement agencies, medicine/healthcare) and costs of breaches (Cisco, 2016), (Cyberedge, 2015), (Enisa, 2017), (European Parliament, 2013), (Gemalto, 2016), (HM Government, 2015), (Identity Theft Resource Center, 2015), (Infowatch, 2016), (Ponemon, 2016), (Trustwave, 2015), (Verizon, 2016), (Verisign, 2016). Concluding from these reports we make specifications about the risk an operational node is opposed to in order to obtain an index ranking which is defined by

$$\text{IND Rank} = \langle P_{(n)}, \text{NoC}, \text{EC}, \text{IND}, \text{Rank}, * \rangle \quad (3)$$

where:

- $P_{(n)}$ is the operational node 1..n
- NoC are the number of connections of a certain type (outgoing, incoming, or both)

- EC is an evaluation criteria: severeness level, attack potential, alteration opportunity which is a multiplication factor
- IND is the confidentiality, availability or integrity index
- Rank = 1..n
- *: EC*NoC = IND

Confidentiality is a characteristic that applies to information. To protect and preserve the confidentiality of information means to ensure that it is not made available or disclosed to unauthorized entities (ISO, 2013). Therefore we take a look at outgoing data flows. Availability is a characteristic that applies to assets. An asset is available if it is accessible and usable when needed by an authorized entity (ISO, 2013). Hence it follows that we must consider incoming data flows. To preserve the integrity of information means to protect the accuracy and completeness of information and the methods that are used to process and manage it (ISO, 2013).

The higher an information is classified, the higher it is and needs to be protected (severeness level) (Rodgers, 2017). Therefore it is easier to attack the availability (attack potential) or alter information (alteration opportunity) at a lower security classification level than at a higher one. On the other hand a high level information is certainly more interesting to be tampered, justifying more efforts to alter and deny access to it (Verizon, 2016). Hence it follows that high level information could be more at risk. From a perspective of negative impacts for the stakeholder the severeness level could like this (classification ≙ multiplication factor):

- TOP SECRET ≙ 6
- SECRET ≙ 5
- CONFIDENTIAL ≙ 4
- RESTRICTED ≙ 3
- UNCLASSIFIED ≙ 2
- PUBLIC ≙ 1

The lower the classification level and its protection the higher the potential of a successful attack to the availability (Infowatch, 2016). All connections could be tampered but it is more likely that information coming from lower sources can be altered more easily due to the fact that the security requirements rise proportionally to the classification level (Rodgers, 2017). The attack potential and the alteration opportunity would be like the following:

- TOP SECRET ≙ 1
- SECRET ≙ 2
- CONFIDENTIAL ≙ 3
- RESTRICTED ≙ 4
- UNCLASSIFIED ≙ 5
- PUBLIC ≙ 6

Table 1: Index Ranking.

Op. Node	Number of CXN (NoC)	EVAL Criteria (EC)	Index (IND)	Rank
P_n	Outgoing	Severe-ness Level	Confi-dentiali-ty Index	1..n
P_n	Incoming	Attack Potential	Availa-bility Index	1..n
P_n	Incoming and Outgoing	Alteration Oppor-tunity	Integri-ty Index	1..n

A table such as Table 1 summarizes these three indexes to identify operational node structures that are most endangered by the loss of confidentiality, availability and integrity. The first column of the table contains the operational nodes $P_{(n)}$ which are the items under analysis. The second column lists the number of connections (NoC) which leave and/or enter the operational node. The third column contains the evaluation criteria (EC) which can be one out if these three: Severeness Level (confidentiality), Attack Potential (availability) and Alteration Opportunity (integrity). The fourth column is the result of column two and three and contains the indexes for the three security aims confidentiality, availability and integrity (EC*NoC=IND). Finally the fifth column shows the ranking (Rank) of the operational nodes which are most endangered.

After identifying a ranking for those three security aims it is now possible to merge these rankings of each operational node depending on the importance of each security aim to obtain a security importance value (SIV) for each node. The weighting lies upon the focus of the modeler and has to be specified by him. If the security aims are weighted equally the equation would be:

$$SIV=W_C*C+W_A*A+W_I*I \quad (4)$$

where $W_C = W_A = W_I = 0,3\bar{3}$ and C = confidentiality ranking, A = availability ranking, I = integrity ranking. The lower the SIV the more endangered the operational node is and possible security measures should be taken on it first.

5 USE CASE

Organizations with certain operational node structures which are dealing with classified data in different security classifications zones define information exchange requirements in order to receive functional demands which a future communication and information system must fulfil.

In this section we analyze a generic example of an operational node structure which will be represented as a DFDsec to show the details of dynamic behavior and to describe the boundaries of the structure as well as their respecting security level. The analysis conducted identifies the operational nodes which are affected most by the threat of losing confidentiality, availability and integrity.

In our exemplary generic use case (Figure 5) there are four elements of the process organization which are processes (P) in the data flow diagram: The Ministry (P1), the Headquarter (P2), the Operations Command (P3) as well as the Team (P4) and two elements outside of the process organization: The Internet (P5) and an Attack Vector (P6).

There are three clusters: DEU_{-1}, DEU_1 and DEU_3. Cluster DEU_{-1} consists of all processes (P1) - Ministry, (P2) - Headquarter, (P3) - Operations Command, (P4) - Team, (P5) - Internet, (P6) - Hacker.

Cluster DEU_1 contains (P1) - Ministry, (P2) - Headquarter, (P3) - Operations Command, (P4) - Team. Cluster DEU_3 includes (P2) - Headquarter, (P3) - Operations Command, (P4) - Team.

Each process which has data flows with the classification DEU_{-1} which is PUBLIC is situated in the security classification zone DEU_{-1}. Each process which has data flows with the classification DEU_1 which is DEU RESTRICTED is situated in the security classification zone DEU_1. Each process which has data flows with the classification DEU_3 which is DEU SECRET is situated in the security classification zone DEU_3.

The Ministry issues strategic directives (F1, F2) to its subordinate offices (Headquarter, Operations Command) which are classified as DEU RESTRICTED and therefore the data flow receive the additional attribute DEU_1. These offices develop plans and concepts (F3, F4 - classified as DEU RESTRICTED) which incorporate regulations for their subordinate units (e.g. Team) or technical/functional concepts which are guidelines for other units. These concepts may have to be approved by the ministry. Sensitive task or reports (F5, F6, F7) sent during a mission are classified as DEU SECRET and obtain the additional attribute DEU_3. Reports sent after a conducted mission (F8) are classified as DEU RESTRICTED in this use case and therefore the data flow receives DEU_1 as additional security attribute. During their work they may need information from the public network (F9, F11, F13, F15). The requested information is PUBLIC (F10, F12, F14, F16). Data flows receive the attribute DEU_{-1}. The Attack Vector represents the focused problem in which someone tries to aspirate classified information and for instance wants to publish it to the public (F17, F19, F21, F23). This can be a malicious insider or an intruder from outside. His data flows are also PUBLIC and receive the attribute DEU_{-1}. If he should receive information back this information will be at least UNCLASSIFIED (F18, F20, F22, F24). In our example it is DEU RESTRICTED. Therefore the data flow receives the attribute DEU_1.

Each process has incoming or outgoing data flows with a security classification. (P1) - Ministry has four incoming and four outgoing data flows. Five of them have the classification DEU_1 and three of them have the classification DEU_{-1}. Hence it follows that (P1) - Ministry was added to the cluster DEU_1 and the cluster DEU_{-1}. (P2) - Headquarter has three incoming and four outgoing data flows. These data flows have three different classifications and therefore (P2) - Headquarter was added to three clusters, DEU_{-1}, DEU_1 and DEU_3. (P3) - Operations Command has five incoming and four outgoing data flows. The data flows have likewise the three different classifications DEU_{-1}, DEU_1 and DEU_3. (P3) - Operations Command was added to three clusters. (P4) - Team has four incoming and four outgoing data flows with the security classification DEU_{-1}, DEU_1 and DEU_3 and were therefore added to three clusters.

In Figure 5 the number of connections have been limited to those connections which are between zones. The information flows within one zone which are flows from zone $I_{(x)}$ to zone $I_{(y)}$, while $x \in Z$, $y=x$ and z = classification level have been omitted, i.e. data flows with attribute DEU_3 (F5, F6, F7).

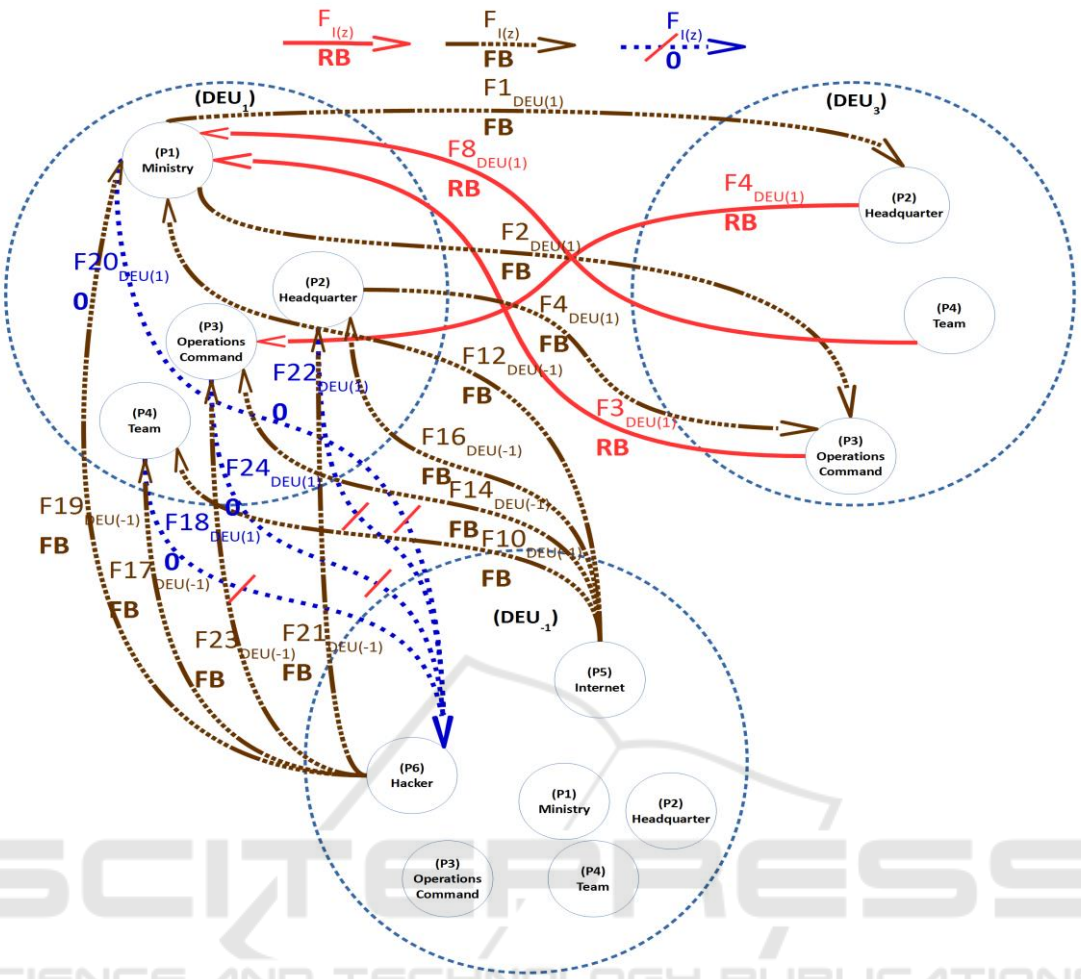

Figure 5: DFDsec.

The information flows between zones need to be looked at closer due to security regulations and the described problem that classified information is threatened from loss of confidentiality, availability and integrity, especially when exchanged between networks with different security classifications. Therefore we highlighted them in different colors (brown, red, blue with red crossing):

- Data flows $F_{I(z)}$ from zone $I_{(x)}$, where $x \in Z$ and z = classification level and $z \leq x$, to a higher security classification zone $I_{(y)}$, where $y > x$ are highlighted in brown and the attribute "FB" = Forbidden Backflow is added. In our use case these flows are (F1), (F2), (F4), (F10), (F12), (F14), (F16), (F17), (F19), (F21) and (F23).
- Data flows $F_{I(z)}$ from zone $I_{(x)}$ to zone $I_{(y)}$, where $x \in Z$, $0 \leq y < x$ and z = classification level and $z \leq y$ are marked in red and the attribute "RB" = Released Backflow is added. In our use case these flows are (F3), (F4), (F8).
- Data flows $F_{I(z)}$ from zone $I_{(x)}$ to zone $I_{(y)}$, where $x \in Z$, $x \geq y$, $y = -1$ and z = classification level and $z > y$ are flagged in blue, crossed out red and the attribute "0" = No Flow is added.

5.1 Analysis

We can now identify those operational nodes which are endangered most of losing confidentiality, availability and integrity. First of all we have a look at the outgoing data flows of each operational node in order to determine the threat of losing confidentiality (see Table 2). (P1) has three outgoing data flows, all of them going from the cluster DEU_1 to the other clusters. The severeness level of disclosing information to an unauthorized entity is one out of three. It can be either 1 which is DEU_{-1} = PUBLIC or it can be 3 which stands for DEU_1 =

RESTRICTED or it is 5 which is defined as DEU_3 = SECRET. Due to the fact that (P1) has only outgoing data flows from cluster DEU_1 the severeness level is 3. Therefore the confidentiality index is the result of three times three which is nine. (P2) has three outgoing data flows, two of them leaving cluster DEU_1 and one of them is leaving cluster DEU_3. The corresponding severeness level is 3 and 5. The confidentiality index calculated from the equation EC*NoC=IND (see section 4) is the result of two times three plus one times five which is eleven. (P3) and (P4) have two outgoing data flows, one of them leaving cluster DEU_1 and the other one is leaving cluster DEU_3. The associated severeness level is 3 and 5. Hence the confidentiality index is one times three plus one times five which is in both cases eight. This implies for a confidentiality ranking that (P2) is threatened most, (P1) is threatened second most and (P3) and (P4) are equally endangered of losing confidentiality to an unauthorized entity.

Table 2: Confidentiality Ranking.

Operational Node	#of outgoing connections (NoC)	Severeness Level (EC)	Confidentiality Index (IND)	Rank
P1	(0/3/0)	(1/3/5)	0+9+0=9	2
P2	(0/2/1)	(1/3/5)	0+6+5=11	1
P3	(0/1/1)	(1/3/5)	0+3+5=8	3
P4	(0/1/1)	(1/3/5)	0+3+5=8	3

The same calculation can be done for the availability and the integrity index. For the availability the incoming data flows of each operational node and the attack potential have to be considered in order to determine the threat of losing availability (see Table 3).

Table 3: Availability Ranking.

Opera-tional Node	# of incoming connections (NoC)	Attack Potential (EC)	Availability Index (IND)	Rank
P1	(2/0/2)	(6/4/2)	12+0+4=16	2
P2	(2/1/0)	(6/4/2)	12+4+0=16	2
P3	(2/2/1)	(6/4/2)	12+8+2=22	1
P4	(2/0/0)	(6/4/2)	12+0+0=12	4

If all data flows of each operational node and the alteration opportunity are regarded we can determine the threat of losing integrity (see Table 4).

Table 4: Integrity Ranking.

Operational Node	# of connections (NoC)	Alteration Opportunity (EC)	Integrity Index (IND)	Rank
P1	(2/3/2)	(6/4/2)	12+12+4=28	1
P2	(2/3/1)	(6/4/2)	12+12+2=26	3
P3	(2/3/2)	(6/4/2)	12+12+4=28	1
P4	(2/1/1)	(6/4/2)	12+4+2=18	4

After identifying a ranking for the three security aims it is now possible to merge these rankings of each operational node depending on the importance of each security aim the security importance value (SIV) of each operational node.

Table 5 shows the SIV ranking. For (P1) and (P3) the SIV is 1.65, for (P2) the SIV is 1.98, for (P5) the SIV is 3.63. Hence it follows that (P1) and (P3) are, under balanced weighting, most endangered.

Table 5: Security Importance Value (SIV) Ranking.

Operational Node	Confidentiality Ranking	Availabilty Ranking	Integrity Ranking	SIV	Rank
P1	2	2	1	1.65	1
P2	1	2	3	1.98	3
P3	3	1	1	1.65	1
P4	3	4	4	3.63	4

We started with an operational node structure and their interconnection data flows. We specified possible (valid and invalid) paths by which information can flow among them. Then we clustered all objects of one security class into one security classification zone and thereupon we reduced and identified all connections to only those connections which flow between these security classification zones. Finally we analyzed the structures with regard to the security aims confidentiality, availability and integrity and resulted, under balanced weighting, that certain operational nodes are endangered most. Therefore security measures should be taken on these first.

6 CONCLUSION

In this paper we presented a security analysis of an organizational structure, which was displayed as a security data flow diagram. This representation enables the identification of security classification

zones, which help to understand allowed and forbidden information flows within and between these zones. We call the resulting model a DFDsec. The model enables a threat analysis on interconnections, especially between the identified security zones, in order to determine operational nodes which are most endangered by the threat of losing confidentiality, availability or integrity. We discussed an initial approach for quantifying the security importance of all nodes, based on the given DFDsec structure. This helps to rank and prioritize operational nodes in their importance for necessary security improvements and mitigation efforts. This approach can be used already in the early phase of the development phase which helps reducing costs.

The DFDsec methodology is work in progress. Future work will focus on the further analysis of structural properties in the data flow representation. We also aim for a quantitative analysis approach, where data flow edges are parametrized with attack potentials. This would allow an even more precise identification of vulnerable operational nodes. Another future topic is the application of the methodology in a practical context, such as the German armed forces IT infrastructure.

REFERENCES

Bell, D. E., LaPadula, L. J., 1976. *Secure computer system: Unified Exposition and Multics Interpretation*, Technical Report ESD-TR-75-306, MITRE Corp. MTR-2997, Bedford, MA.

Biba, K. J., 1976. *Integrity considerations for secure computer systems*, Technical Report ESD-TR-76-372, MITRE Corp. MTR-3153, Bedford, MA.

Cisco, 2016. *Cisco 2016 Annual Security Report.*

Cyberedge Group, 2015. *2015 Cyberthreat Defense Report.*

DeMarco, 1978. *T. Structured Analysis and System Specification*, Yourdon Press, New York, NY.

Denning, D. E., 1976. *A lattice model of secure information flow*, Communications of the ACM, 19(5):236-243.

Denning, D. E. and Denning, P. J., 1977. *Certification of programs for secure information flow*, Communications of the ACM, 20(7):504-513.

Eckstein,C., 2015. *Preventing data leakage: A risk based approach for controlled use of the use of administrative and access privileges*, White Paper, SANS Institute.

ENISA, 2017. *ENISA Threat Landscape Report 2016*.

European Parliament, Directorate General for Internal Policies, Police Department A: Economic and Scientific Policy, 2013. *Data and Security Breaches and Cyber-Security Strategies in the EU and its International Counterparts.*

Federal Office for Information Security (BSI), 2008. *Information Security Management Systems (ISMS), BSI-Standard 100-1*, Version 1.5.

Gane, C. and Sarson, T. 1977. *Structured Systems Analysis and Design*, Improved Systems Technologies, Inc., New York, NY.

Gemalto, 2016. *Breach Level Index*.

Genua gmbh, 2016. *Datendiode vs-diode*, Munich, Germany: www.genua.de, Web-Access 06. April.

Gordon, P. 2007. *Data Leakage♁Threats and Mitigation*, White Paper, SANS Institute.

Harrison, W. S., Hanebutte, N., Oman, P. W. and Alves-Foss, J., 2005. *The MILS Architecture for a Secure Global Information Grid*, The Journal of Defense Software Engineering, pages 20-24.

HM Government, 2015. *2015 Information Security Breaches Survey*.

Identity Theft Resource Center, 2015. *2015 Data Breach Stats*.

Infodas, 2016. *SDoT ® Security Gateway 5.0*, Cologne, Germany: www.infodas.de, Web-Access 06. April.

Infowatch, 2016. *Global Data Leakage Report, H1 2016*.

International Organization for Standardization, 2013. *ISO 27001 Information technology - Security techniques - Information security management systems Overview and vocabulary*, GE, SUI.

Mah, P. 2017. *7 Social Engineering Scams and How to Avoid Them*, www.cio.com, Web-Access 09. June.

Myers, A. C. and Liskov, B., 1997. *A decentralized model for information flow control*, In SOSP 97: Proceedings of the sixteenth ACM symposium on Operating systems principles, pages 129142. ACM Press.

Ponemon Institute, 2016. *2016 Cost of Data Breach Study: Global Analysis.*

Rodgers, C. 2017. *Data Classification: Why is it important for Information Security?*, SecureState Blog: www.securestate.com, Web-Access 05. July.

Rushby, J. 1981. *Design and Verification of Secure Systems*, ACM Operating Systems Review Vol. 15 No. 5 pages 12-21. ACM Press.

Schmidt, K., Tröger, P., Kroll, H., Bünger, T. et al., 2014. *Adapted Development Process for Security in Networked Automotive Systems*, SAE Int. J. Passeng. Cars Electron. Electr. Syst. 7(2):516-526, doi:10.4271/2014-01-0334.

Shostack, A., 2008. *Experiences Threat Modeling at Microsoft, In Workshop on Modeling Security*, Toulouse.

Torr, P. 2005. *Demystifying the Threat-Modeling Process*, IEEE Security & Privacy Magazine, vol. 3, no. 5, pp. 66-70.

TrendLabs SM APT Research Team, 2012. Spear-Phishing E-Mail: die beliebteste APT-Angriffstechnik}, Trend Micro, Hallbergmoos, Germany.

Trustwave, 2015, *2015 Trustwave Global Security Report*.

Verisign, 2016. *Verisign Distributed Denial of Service Trends Report*, Volume 3, Issue 3.

Verizon, 2016. *2016 Data Breach Investigations Report.*

Yourdon, E. 1989. *Modern Structured Analysis*, Yourdon Press, Upper Saddle River, NJ.