

# Ftp server in a data diode

Rusu George, Boulif Ilias, Orinx Cédric

November 16, 2017

## **Contents**



Figure 1: A diode schema in electronics.



Figure 2: The schema of the usage of a data diode in a company network.

## 1 Concept of operations

### 1.1 Introduction to the problem

Our client would want to prevent his research and development labs against industrial espionage and confidential information leaking. He wants to divide the company network in order to maintain the labs in an isolated environment. Thus, the critical data could not leave the company using the network. One problem related is assuring that all the operating systems are up-to-date. This could be achieved using a manual operation where an employee would manually update every node in the isolated network using a mass-storage device<sup>1</sup>. However, this is not the most optimized way and for sure it is not cost less and timeless for the company, without mentioning that there could be dependency problems. This method is also prone to human errors and this could generate security vulnerabilities.

### 1.2 Our solution

In order to address this problem we are going to implement a data diode. In electronics as shown in Figure 1, a diode is a component which conduct the current in one direction. Thus the term of data diode is a set of components that only let the data to travel in an unidirectional way. An example of use case is illustrated in Figure 2, letting the data pass from a low risk LAN to a high risk LAN but not the way around.

A data diode is made using one transmitter(Tx) and one receiver(Rx) both linked using only one fiber cable. Thus, when a digital data is sent through, every bit will be converted<sup>2</sup> into an electrical pulse which will be convert in light pulse using a LED<sup>3</sup>. Then, the output light will pass trough the fiber cable and as soon as the photons are reaching the receiver, they will be converted firstly into an electrical pulse by a photo diode and then in bits<sup>4</sup>. Since there is only one fiber, data can only travel in one way.

One major problem of this system is the network transport protocols(layer 4 in the OSI model) that applications are using. The most used transportation protocol nowadays is TCP. However, TCP is a bi-directional protocol which need the 3 way handshake in order to start a connection, thus there must be at least two fiber cables. In contradiction, UDP does not require a handshake process because it does not provide reliability, ordering, data integrity and does not set up a dedicated end-to-end connection automatically. Thus, UDP protocol does not necessary require a bi-directional data transfer. Therefore, the use of such a protocol is recommended when dealing with applications that are not sensitive to data

<sup>1</sup>A USB-key for instance.

<sup>2</sup>Using a Digital-to-analog converter (DAC).

<sup>3</sup>Light Emitting Diode.

<sup>4</sup>Using an analog to digital converter (ADC).

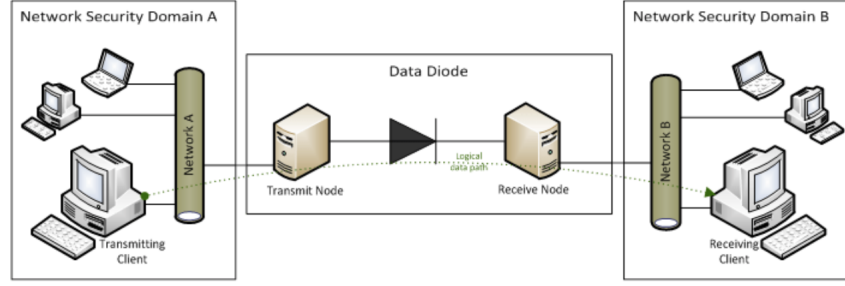


Figure 3: Data logical path within the data diode.

loss or that implements an error checking system. Hence, the use of the UDP protocol within the data diode.

It seems obvious now that the transport protocol used between the transmitter and the receiver of the data diode is UDP. However, this brings multiple problems. One of the problems is how to be able to use applications that requires TCP. Another problem could be the reliability of the digital data transfer in the data diode, how can we ensure that there will not be any data loss during the transfer. We will address to those questions further down in this document.

Furthermore, system administrators would dispose of a file transmission mechanism from the low level to the high level in order to maintain all the computers up to date. This will be done using the FTP protocol and will be explained in the next section.

### 1.3 Our System description

The data diode concept is well orchestrated combination between hardware and software. The data diode is composed of two servers: a transmitter and a receiver. We could consider that the transmitter server is connected to the low security risk LAN where a connection to the internet is made and where all the employees are connected. The receiver is connected to the high security risk LAN where all the labs are connected and where critical informations are stored.

The transmission of the information will be from the low side to the high side and blocked the way around as shown in Figure 3.

#### 1.3.1 Hardware

The following table presents the components used for the data diode implementation. We should mention that both transmitter and receiver possess two NICs<sup>5</sup>: one for the communication within the LAN and one for the data diode.

Components	Description
Transmitter server	a physical machine or a virtual machine. The NIC for the data diode will be set in transmission mode only.
Receiver server	a physical machine or a virtual machine. Similar to the transmission, the nic will only be set in a receiver mode.
One fiber/UTP	This optical fiber allows communication between the low server the high server. It will probably be simulated using a UTP/RJ45 cable.
4 Fiber NIC/UTP NIC	Two for each server (transmission and receiver).

Table 1: Components description

<sup>5</sup>Network interface controller.



Figure 4: Entire system.

### 1.3.2 Software

The Linux distribution which will be used in our project will be GNU Debian. We will have two linux machine, one on the transmitter side and one on the receiver side.

The data diode will be easily administrated using a web interface running on an Apache server. The web site will be hosted on the transmission server. It will be described in more details further in this document.

Every configuration script used by the web interface will be a combination of Python language and Bash language.

### 1.3.3 First overview on the project implementation

Let's start by explaining which protocol shall be used in the data diode. Remember that our main goal is to create an isolated environment but also to facilitate the update process of the machines within the high security zone. As we mentioned earlier, we are going to use the FTP<sup>6</sup> protocol. There are some alternatives to FTP such as BlindFTP, SFTP, etc. However FTP and SFTP are using TCP as their transportation layer protocol. Thus, we are going to use between the transmitter and the receiver the BlindFTP which use UDP instead.

BlindFTP is a simple Python script which was especially created for the communication between the two servers over an unidirectional network. It is simply a tool for transferring files from one side to the other. Despite the fact that use UDP, there is no acknowledgement of received packets but this is compensated by a redundancy of the data. An other advantage is the language used, as it is written in Python the result code is relatively simple and very portable to Windows or MacOS.

It is now time to have a look to our entire system, let's have a look in Figure 4.

Using BlindFTP allows us to ignore if the application is using TCP or UDP, the only thing we need is to make sure that the data arrive to the transmitter. We should note that between the transmitter and the other peers a TCP connection can be used to communicate. And this is the same for the receiver, after the data goes out from the network diode the transportation protocol could be TCP with other computers. In this way, we are not dealing any more with the problem of application transportation layer.

## 1.4 Users

In general every employee of a company has a unique ID, a grade, a job title, etc. The company owner or the manager in charge with the security could choose which grade or which person can have access to the administration web interface, of course this exclude the systems administrators. Thus to make our explanation simple, let's suppose that there will be two types of users in our system : the *administrators* and the all the other *users*.

---

<sup>6</sup>File transfer protocol.

A *user* is simply an employee of the company. Every employee have limited access into the company's files according to their job title or to the specific rule defined by their supervisor.

An *administrator* is a user in charge of operating and maintaining the data diode. It is the only authorized person to interact with the data diode and this is mandatory. Because he will have to know how the data diode is working, should know how to interact with it using the web interface but mostly he should be able to reconfigure it or restore it as quickly as possible in case of a security breach or a system failure. However, an administrator could not have access to some strict secret files of the company but only to the used IT mechanism.

Every user information is managed and stored by the company in their own databases.

## 1.5 User Interface

As we mentioned earlier, the administration of our data diode will happen through a web interface. There will be one web interface hosted on the transmitter which will allows to transmit files or update packages from the lower network to the higher network. But, on the other hand there will also be a receiver web interface for each user to manipulate his transferred data more easily.

Each user should login from both side to use the platform. Here is an overview of our login page:



Figure 5: Login page.

Once login is made, the user will be prompted to the web interface of the transmitter or the receiver. The page are shown in Figure 6 and Figure 7.

On the other hand, an admin can transfer files but also can transfer updated packages from official sources or he can upload a package from his own PC. The admin interface is shown in Figure 8 and in Figure 9.

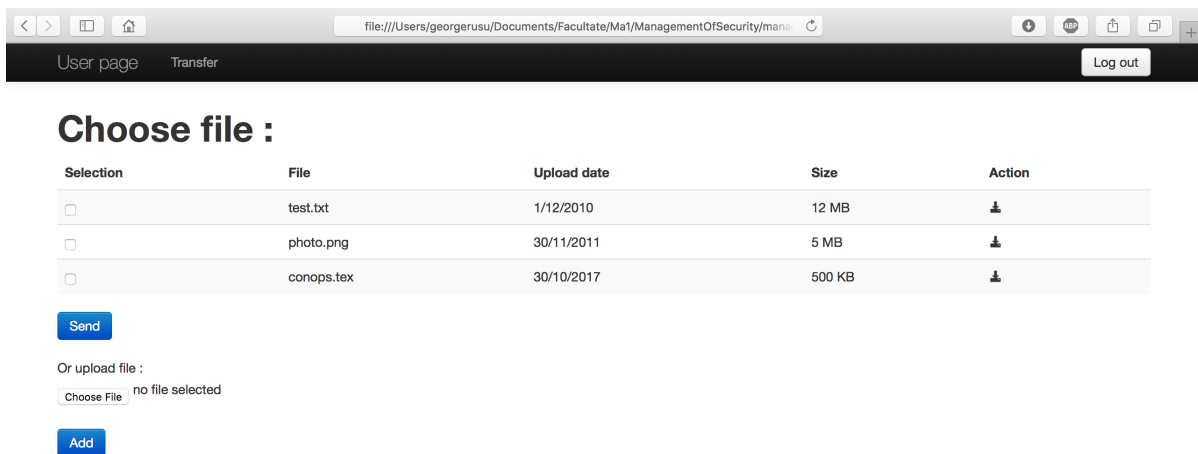


Figure 6: Transmitter user page.



Figure 7: Receiver user page.

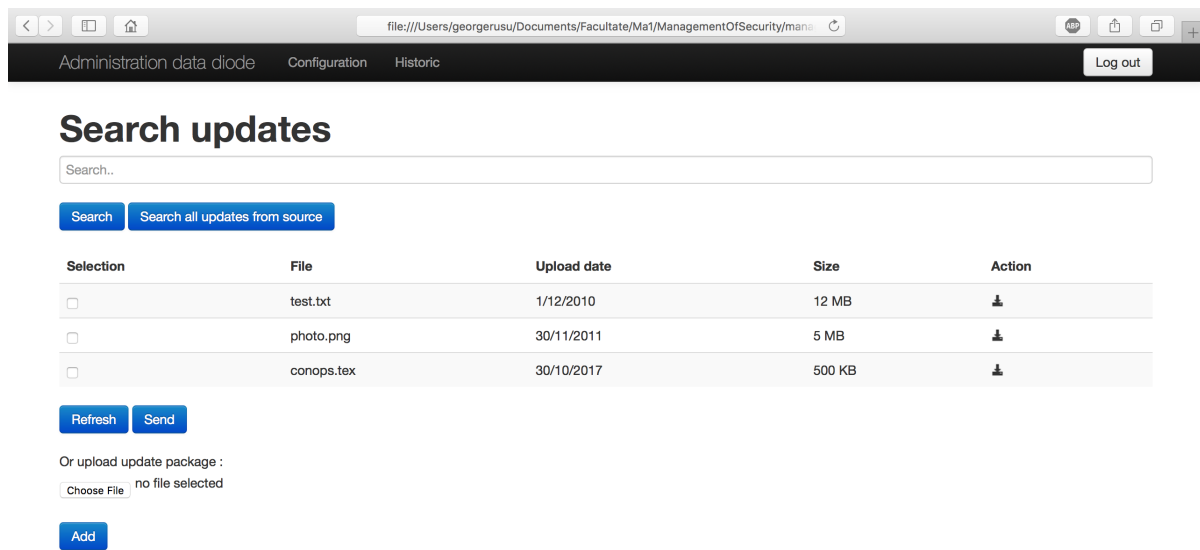


Figure 8: Transmitter admin page.

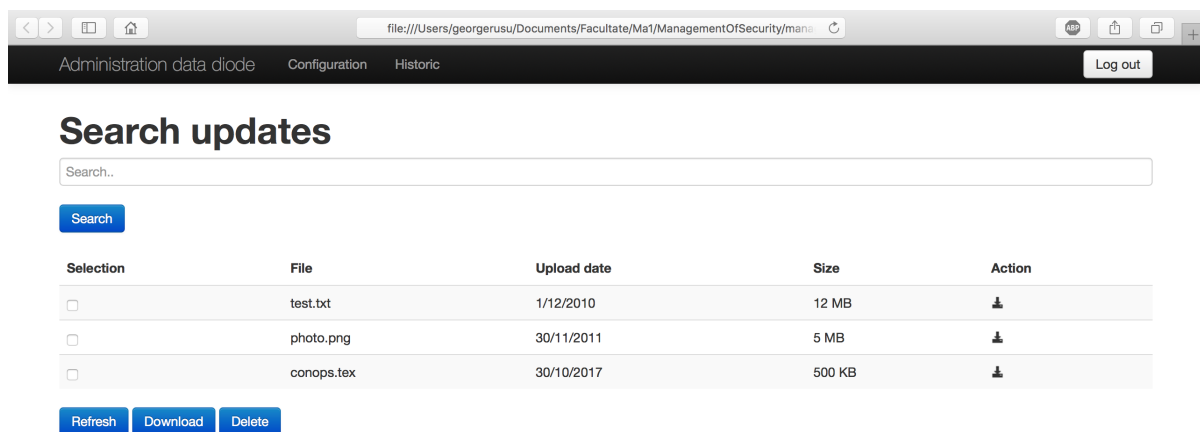


Figure 9: Receiver admin page.



## 1.6 Usage of the WEB UI

In this section we will explain how to use our web interfaces from different point of views.

### 1.6.1 General usage

In both cases, all users have to identify themselves through the login page with their credentials. They must fill in the form with their own username and password (Figure 5). Once the form is submitted, the authentication will bring them respectively to their dedicated page.

In the case of a user connection, he will be redirected to the user transmission page according to his network location: if he is in the low security zone he will be prompted to use the transmission interface (Figure 6) or he will get the receiver interface page (Figure 7) if he stand on the high security zone (after the data diode).

On the transmission page the user can select the files he needs and send them through the data diode by using the *send* button. The user have also the possibility to upload a file from a computer and to add it in the selection list in order to transfer it. On the reception page, the user has a list of all the received files which he can depending on the button chosen and his selection, refresh, download or delete.

When an administrator will log in, similarly to the user mechanism, he will be redirected according to his network location : on the transmission interface or on the receiver interface, respectively Figure 8 and Figure 9.

On the transmission page, a *searchbar* is displayed to allow the administrator to search for available updates. He can search a specific update (button search) or display all available updates from official repositories(button search all updates from sources). In this way he can select from a list the updates that he wants to transfer to the secure network. He can also upload files or new update packages from his computer.

Furthermore, the administrator has an history page. With this page he can browse all the updates installed previously in the system but he can also monitor all the passing data traffic between the two servers. This is made possible using a log information (user, date, size,...).

On the reception page, the administrator has access to all updates that have been send through the data diode. He can update all the machines using some Linux script or to use the web interface to download each update package manually on the desired machine and install them manually. Of course this manual method is not the most optimal but it could serve in case of a need for a system downgrade or an application downgrade on one workstation.

The administrators have the rights to manage all the updates within FTP sharing systems, they can choose to keep or delete updates.

### 1.6.2 A Technical file transfer usage

We have said earlier that there will be a web interface running on both side of the data diode. When a user will want to send a file, he would have to first login. The login page will use the data base of the company in order to authenticate users and to verify each one permissions. However, on the high security zone, the database would not be reachable because of the unidirectional data diode. Thus, once a user would want to send a file to the other side, there will be a daemon which will create a permission and ownership system in order for the other side to recognize to whom the file belongs to. For example, when the UserX would want to send the *file.txt* to the other side, the transmitter daemon will rename the file using the following rule:

$$file.txt \triangleright userName\_password\_fileName\_timestamp\_size\_extension.zip$$

Every file will be converted into a zip archive in order to reduce the initial size, thing that could sometimes have a major impact on the speed of the transfer (hence the time).

On the other side, there will be of course another receiver daemon which will for each file received, decompress the file into the original extension<sup>7</sup>. However to ensure that only the owner (thus the user-Name) may see the file, the second daemon will create for every userName a folder where received files will be stored. The rule for this operations is:

---

<sup>7</sup>This is why the initial extension is putted in to the name of the sent file.

*userName\_password\_fileName\_timestamp\_size\_extension.zip* ▸ folder: *userName\_password* ▸ files in folder: *file.txt*

Furthermore, when a user wants to login from the secure zone of the data diode, instead of queering a database, the web interface will query a daemon which will use all the folders name to authenticate the user. Every password will not be shared in clear text, a hash of SHA-256 will be used. In this way there is no need of database duplication. The main verification will take place on the transmitter side where the database can be accessed: if a user has the permission to transfer a file, he will be able to transfer the file from the transmitter side. Thus, if the file is transmitted to the receiver side, it means that the user has an access granted on the security zone and thus he has already been verified. However, if a user has not yet used the transmission system of the data diode, he will not be able to connect to the interface from the receiver side.

### 1.6.3 Configuration panel for the admin

The admin can configure the IP address of both the transmitter and the receiver. In this case, we are not talking about the communication within the data diode but about the communication to other peers of the network. Using the following panel, administrator can manually configure the static IP for the servers hosting both the web interface and the data diode.

Administration data diode Configuration Historic Log out

## Set server IP

Websocket client

IP address :  .  .  .

Subnet mask :  .  .  .

Gateway address :  .  .  .

Apply changes Restart the server

ATTENTION! In order to apply changes a restart of the system is required.

Figure 10: Configuration panel for admin.

## 2 Risk Analysis

### 2.1 Introduction to the risk analysis

In this section, we are going to assess any risk that could harm or lead to the destruction of our data diode. But first, let's define what is a risk. Generally the term risk is associated with a damage or a loss after an event. Thus we can illustrate the risk by the following rule:

$$Risk(e) = impact(e) \times likelihood(e), \forall e \in Events$$

In our risk analysis we will start by defining all the assets of our project. Those assets could be threatened by some hackers which could exploit some vulnerabilities found and harm the entire system. Hence, the need to assess all potential threat sources and all the vulnerabilities and their impact. Finally, we will conclude by presenting some countermeasures.

#### 2.1.1 Physical Assets

An Asset can be physical such as computers or printers. In our client work space, we could have the following components categorized as physical assets:

- All the computers of the low security zone. Those computers could be employee's personal workstations such as laptops but also company's desktop workstations.
- All the company's firewall systems (routers and switch). Despite the fact that they are not part of our implementation, they could represent a potential asset for malicious persons.
- Our data diode implementation. This comprises our transmitter server and our receiver server.
- Data diode web servers for the web user interface.
- The company data base server.
- All the computers within the high security zone.
- All the servers of the company such as Mail servers, FTP servers and other services servers. This comprises the servers in low and high risk networks.

To all those components a stakeholder would want to have physical access in order to harm the well being of our client company. Thus, it is very important that the physical integrity of every equipment shall be preserved. In other words, a physical access to some physical assets such as routers or servers could harm or destabilize the company core system, this could easily be done for example by simply cutting the power source cable or by disconnecting the internet cable of a server (cutting the communications).

#### 2.1.2 Logical Assets

A logical asset could be a sensitive information. A list of logical assets used by our client's company could be:

- The company data base containing sensitive information.
- The data diode web interface.
- Data on every employee personal computer or desktop workstation used for company purpose.
- Data within the research labs.
- All the hard drives data of every computer company.
- Company's storage system.
- Company's operating systems: this contains also every application used within the company. As an example, Microsoft Office could be used by the employee of the finance department.

In other words, logical assets represent all the sensitive information which belongs to the company and all the software and applications used by the company's employees.

### 2.1.3 Persons involved into the company

At this point, we should identify every person in the company which could have access to some physical and logical assets. Generally in a company, there are several departments in which employees or contractors are working. In order to be able to distinguish every user, we should tag every user with a value according to his loyalty and his hard work for the company. Let's fix 4 valuable stage:

- Very loyal and very hardworking employee.
- Loyal and hardworking employee.
- Not loyal and not hardworking employee.
- Contractors.

It should seem obvious that a very loyal employee which invests a lot of time and work is unlikely to want the destruction or the malfunctioning of the company, because he knows that he will probably lose his job. On the other hand, a non-loyal employee or a contractor could easily want to destabilize the company activities for any kind of reasons<sup>8</sup>.

However, as we do not know every employee of our client's company, we will simplify the picture by mainly focusing on the administrators and the other users. Thus, there are two categories of person :

- The administrators which could have a disastrous impact, they have access to both logical assets and physical assets.
- All other users which could have a moderate impact, they only could have access to physical assets.

Thus, identifying employees using a value tag could be very useful when identifying potential threat sources.

## 2.2 Threat Sources

As we mentioned earlier, identifying potential threat sources can be a difficult operation. However, a system failure could be caused by inevitable events or in contrary, it could be caused by a lack of interest. In order to have a clear view over these kinds of events, we will present all those events using two categories:

- Unpredictable or inevitable events.
- Predictable or avoidable events.

### 2.2.1 Unpredictable events

Such events could be related to natural disaster which cannot be always predicted in advance. Thus, there is no way in which to prevent the mass destruction but the company could take measure in order to minimize the impact. This could be related to natural disaster like:

- Earthquakes
- Storms
- Floods
- Sinkholes.
- Volcanic eruptions
- Tsunami
- Wildfires

However, some event could have a much higher probability of appearance depending on the location where the company has her headquarters. For example, it is more likely to have floods or a tsunami if the company is located near the equator of the globe or in a tropical area. Hence, the importance of the location chosen. For example, the company could prevent some of those, for example choosing where to host their main servers.

---

<sup>8</sup>An employee could receive a job offer from a competitor or a contractor who has not been paid for his work.

### 2.2.2 Predictable events

We can define predictable events as security lack that we are or not acknowledged about our system. Here is a list with some natural person which could be implicated in the hacking action of our client's company system and thus hacking our data diode implementation.

- Employees or company daily users
- Administrators
- Visitors
- Script Kiddies
- Skilled Hacker

Each one in this list could be a potential threat source or serve as so. Let's assume the following scenarios in which the system could get infected.

After a lot of work and time spent on projects for the company, an administrator or simply an employee could for example make a mistake which could destabilize the entire system. As an example, we could assume that an administrator forget to close a configuration port or an employee could share without his willing his login password. As we said earlier, tagging every user of the company using a value could be a good thing when assessing the threat source. Indeed, a valuable employee despite his tiredness would for sure look twice before finishing his work or sharing some sensitive information of the company. But this is not the same for a newbie which just started on his new job.

Furthermore, having visitor into the company can be a good thing for the image and the brand, however it could be very dangerous. Assuming that a visitor could install a *Femtocell* into the company which could get him full access into the network. Despite the fact that it would be for sure the low security network, the malicious user could sniff every network packet transiting trough and thus could steal some information.

For the last two categories, they represent a constant threat for every company. Hence the need for a security policy always up to date and a good strategic network and workstation deployment in order to minimize the risk.

We will detail more about all of this by explaining more vulnerable scenario in the following sections.

## 2.3 Vulnerabilities

In this section, we will give a set of possible vulnerabilities that we should take into account in order to secure the most possible our data diode. In order to have a clear view, we will explain every vulnerability by ordering them according the assets type affected.

### 2.3.1 Affectation je suis ici :)

### 2.3.2 Vulnerabilities Affecting Physical Assets

### 2.3.3 Vulnerabilities Affecting Logical Assets

### 2.3.4 Vulnerabilities Affecting Persons

### 2.3.5 Vulnerabilities Affecting Intangible Goods

## 2.4 Risks and Countermeasures

The following table will give the impact that each vulnerability could have, with a brief description of the impact and its level.

Now we define the likelihood of each vulnerability and describe why such a level is given.

After having analysed the impact and the likelihood of each vulnerabilities, we can define for each of them a risk level using the Table ?? to adapt our countermeasure accordingly.

### 2.4.1 Countermeasure

ID	Vulnerability	Source(s)	Affecting	Description
1	Power outage	Nature	Availability	In the case of a power outage, obviously the data diode system won't work because it needs electrical power in order to run. Hence, availability is compromised.
2	Fire	Nature	Availability	A fire in the building which is in direct contact with the data diode or because of the heat will cause damage to the data diode system and this will lead to his non-operation and thus non-availability.
3	DDOS attack	Skilled Hacker	Availability	A DDOS is a cyber-attack where the hackers seeks to make a system unavailable to its users by temporarily or indefinitely disrupting services of a host connected to the Internet. The attack is accomplished by flooding the targeted resource with superfluous requests from many different sources in an attempt to overload systems and prevent some or all legitimate requests from being fulfilled. The administration of our data diode happens through a web interface which means that it is vulnerable to DDOS attack. In case of DDOS attack the data diode will become unavailable as long as the attack lasts.
4	Data diode hardware failure	Nature	Availability	If a failure in the data diode's hardware occurs, such as the fiber cable, it can lead to the non-operation of the data diode.
5	Zero Day Attack	Skilled Hacker	Availability, Confidentiality, Integrity	Since there is possibility that a new attack is created, this attack could be a vulnerability due to the fact that we still don't know how to protect ourselves. Since then, it could affect one or more part of the information security.
6	Data diode hardware degradation	Unauthorized user	Availability, Confidentiality, Integrity	This vulnerability is possible once an unauthorized person has access to the data diode itself. This could lead to a simple destruction of the data diode or robbery of some element (availability), trying to remove data from it (confidentiality), or trying to modify/add component to falsify the data (integrity). There is also a possibility that the degradation is done without bad intention (disconnecting a cable, spilling a drink on it).
7	SQL injection	Skilled Hacker, script kiddies, unauthorized user		
8	Brute force attack			
9	Virus and malwares			
10	Social engineering			
11	Misconfiguration			
12	Data credentials lost			
13	Data credentials theft			
14	Failure in maintenance from the provider			
15	Remote connec-			

ID	Vulnerability	Impact	Description
1	Power outage	Medium	The data diode will not work anymore which means it will not be available until the power will be re-establish. A power outage usually does not exceed one hour which mean there is not a high impact.
2	Fire	High	The impact depends on the damages caused by the fire. The data diode can suffer significant damages which could lead to the replacement of the hardware.
3	DDOS attack	High	If the website of the data diode suffers from a DDOS attack, the secured network productivity can grid to a total halt if the transfer of important files cannot be done. This can lead to an important cost for the business, it is why a DDOS attack can have a big impact.
4	Data diode hardware failure	Low	The data diode hardware are well chosen and tested in order to keep a low risk of failure.
5	Zero Day attack	High	Since we don't know the attack, its impact could be as low as high.
6	Data diode hardware degradation	High	This vulnerability could affect the three part of the information security and also lead to non-reversible damage.
7	SQL injection		
8	Brute force attack		
9	Virus and malwares		
10	Social engineering		
11	Misconfiguration		
12	Data credentials lost		
13	Data credentials theft		
14	Failure in maintenance from the provider		
15	Remote connection		
16			

Table 3: Impact

ID	Vulnerability	Likelihood	Description
1	Power outage	Low	Weather is responsible for the majority of major power outages that occur, but there many others causes such as short circuits, blackouts and so on. Thus power outage has a medium likelihood to happen but in the case where the client's company is equipped with an <b>uninterruptible power supply (UPS)</b> , which is an electrical apparatus that provides emergency power when the mains power fails, the likelihood becomes low.
2	Fire	Low	A fire can occur due to an electrical incidents, accidents or sabotages caused by human actions.
3	DDOS attack	High	If the company is a popular, the secured network will be a prime target for the hackers. therefore, attempts may be numerous. Furthermore, DDOS attacks have grown in scale. DDoS attacks are popular because in some ways, they're easy to do. So even a beginner hacker may be able to make an attack.
4	Data diode hardware failure	Medium	In the case of hardware failure within the data diode, it is the whole operation of the system that can be disturbed.
5	Zero Day attack	Medium	This kind of attack is constantly sought after by different people like skilled hacker, but its difficult to find one, and more of this, find one that can affect our system.
6	data diode hardware degradation	Medium	
7	SQL injection		
8	Brute force attack		
9	Virus and malwares		
10	Social engineering		
11	Misconfiguration		
12	Data credentials lost		
13	Data credentials theft		
14	Failure in maintenance from the provider		
15	Remote connection		
16			

Table 4: Likelihood

Risk Level			
	Impact		
Likelihood	Low	Medium	High
Low	Low	Low	Low
Medium	Low	Medium	Medium
High	Low	Medium	High

Table 5: Risk Level



ID	Vulnerability	Source(s)	Countermeasure(s)	I	L	Risk Level
1	Power outage	Nature	For this vulnerability we accept the possible risks.	M	L	Low
2	Fire	Nature	The data diode must be placed in a room close by a fire door to mitigate as much as possible the risk of the data diode being affected by fire.	H	L	Low
3	DDOS attack	Skilled hacker		H	H	High
4	Data diode hardware failure	Nature	Since an administrator must be ready to intervene in case of a failure of the hardware, we can accept the risk of this vulnerability.	L	L	Low
5	Zero Day attack	Skilled hacker	Due to the fact that we can't be prepare against this attack, we accept the risks of this vulnerability.	H	M	Medium
6	Data diode hardware degradation	Unauthorized user	The data diode must be place in a closed room lock with a key or access card possessed only by the administrator. A camera coul also be placed at the entrance to the room. Since the only person who can access to the room is the administrator, the risk is limited to a clumsiness of the administrator.	H	M	Medium
7	SQL injection					
8	Brute force attack					
9	Virus and malwares					
10	Social engineering					
11	Misconfiguration					
12	Data credentials lost					
13	Data credentials theft					
14	Failure in maintenance from the provider					
15	Remote connection					
16						

Table 6: Countermeasure

## References

- [1] DEEP SECURE, *How Does a Data Diode Work?* Discussion Paper, February 2017.
- [2] SANS Institute InfoSec Reading Room, *Tactical Data Diodes in Industrial Automation and Control Systems*, January 2015.
- [3] CS Risk Management and Compliance Ltd, *Data Diode vs Firewall Feasibility*, September 2016. [Online]. Available: <https://csriskmanagement.co.uk/data-diode-vs-firewall-feasibility/> [Accessed: 30- Oct- 2017]
- [4] BlindFTP, *Blind ftp protocol*[Online]. Available: <https://www.decalage.info/fr/python/blindftp> [Accessed: 30- Oct- 2017]