symmetry

MDPI

Article

Web-Browsing Application Using Web Scraping Technology in Korean Network Separation Application

Won-Chi Jung [1], Jinsu Kim [1] and Namje Park [2,*]

[1] Department of Convergence Information Security, Graduate School, Jeju National University, Jeju 63293, Korea; jwonchi@jdcenter.com (W.-C.J.); kimjinsu@jejunu.ac.kr (J.K.)
[2] Department of Computer Education, Teachar's College, Jeju National University, Jeju 63293, Korea
* Correspondence: namjepark@jejunu.ac.kr

Abstract: Attackers' intrusion into the Enterprise LAN is increasing every year, and the method is becoming more intelligent and crafty. Various security measures against external network intrusions, such as firewalls, are being studied and applied to protect against external attacks, but it is difficult to respond to increasing attacks. Most institutions block access from the external network for the safety of the internal network and allow access from the internal network to the external network through some restricted ports. In particular, restricted ports in subject to a variety of security techniques to block intrusion into the internal network, but in the process, access to the internal network is only applied by restricted ports, making it inefficient to handle internal requests. Although various studies have been conducted on network isolation to address these challenges, it is difficult to perform tasks efficiently as security functions, such as detecting whether request data is attacked or not, during actual application. The proposed technique is a network-blocking-based network separation technique that converts data from the external network connected to the Internet into symmetry data from which malicious code is removed through an agent and delivers it to the client of the internal network. We propose a technique to provide access.

Keywords: network separation; network separation policy; web scraping; web crawl; headless browser

Citation: Jung, W.-C.; Kim, J.; Park, N. Web-Browsing Application Using Web Scraping Technology in Korean Network Separation Application. *Symmetry* **2021**, *13*, 1550. https://doi.org/10.3390/sym13081550

Academic Editors: Nikos Mastorakis and Juan Luis García Guirao

Received: 27 July 2021
Accepted: 20 August 2021
Published: 23 August 2021

1. Introduction

The utmost threats to cyber security in this era coexist with countermeasures to deal with them. In Korea, the network separation policy was adopted in 2007 as a countermeasure against cyber-attacks from overseas against state agencies [1]. All the state agencies, government ministries, local governments, their affiliated agencies, and public institutions were included in the scope of application. In the beginning, our government stuck to physical separation, but, in time, logical separation was allowed. In 2013, it mandated the network separation of information and communication service providers and Internet service providers, and, in 2014, including the financial sector (banks, insurers, securities, and credit card companies). As such, network separation has been applied to the majority of institutions such as government agencies, public institutions, and the financial sector [2].

In the case of a network separation environment, the more the security is strengthened, the more processes to go through to acquire data from the external network. Such an increase in security level increases the time required and leads to a decrease in work efficiency. Especially in Korea, the government announced a regulation in 2020 calling for the mandatory introduction of network separation for fintech companies concerning fraudulent payments using a user's personal information in the simple payment system of a financial company. Fintech-related companies may now be obligated to comply with strict network separation regulations, and a significant reduction in efficiency is expected [3,4].

This possible reduction in efficiency caused by network isolation regulations needs to be minimized and requires more efficient implementation of requests using limited external network access paths. In general, when data are requested from a website of an

external network in the Enterprise LAN, it occurs as a link is performed from a hyperlink or a sentence executed by JavaScript to a path with an attack intention. These issues can be effectively countered by removing the material's attack script.

In this paper, we propose the methods to improve efficiency while maintaining the security policy in a network isolation environment: the means to transform the web to non-malicious via web scraping technology, that transforms a web page into a PDF or image form rather than the web to make it content disarm and reconstruction (CDR). In this proposed method, the agent, which is connected to the external network, removes the malicious code by using the web scraping technology for the data requested by the client. Afterwards, the agent can reduce the possibility of infection of the internal network by removing the malicious code and then transmitting the data that is completely symmetry with the data requested by the client.

2. Related Works

2.1. Science of Security (SoS)

Rajendra et al. proposed a hypervisor-level distributed network that identifies feasible network traffic features and accurately detects attacks, arguing that current security solutions, such as firewalls, intrusion detection and protection systems, need to be extended to solve these issues [5]. Figure 1 shows the stability of network isolation, which shows that physically independent servers are logically isolated through virtualization on each server, which shows that the stability of network isolation can be applied in a logical network environment, as well as a physical server.

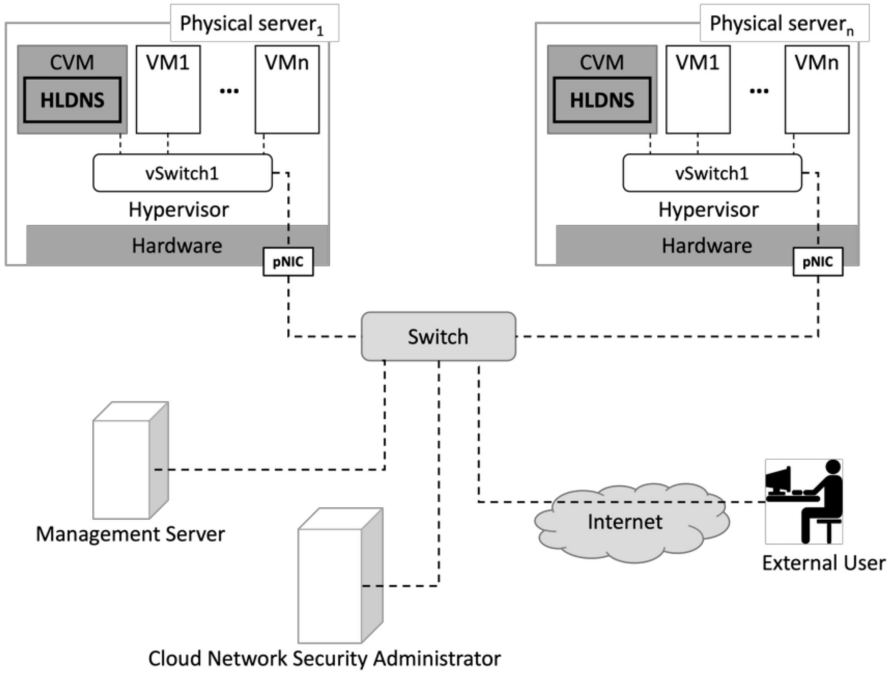


Figure 1. HLDNS framework in cloud infrastructure.

2.2. Security Assessment Model for Network Separation

The DongHwi study [6] provided three perspectives, which are defined as a criterion by system level, a criterion by network equipment level, and a criterion by security equipment level, on the criteria for evaluating vulnerabilities by identifying each component's weaknesses in a network separation environment and analyzing possible threats and impacts [7]. They minutely defined a model that evaluates the level of integration for the most influential network section. As can be seen in Table 1, this is used as a basis for developing a cybersecurity control strategy depending on the effects of network separation.

Table 1. Critical network segment evaluation level.

Level	Value	Detail
L5	Very High	It can cause great damage to the internal network, Causes fatal side effects to server operation
L4	High	It can cause damage to the internal network, Causes side effects to server operation
L3	Moderate	It can cause great damage to the external network, The service may be restricted, Damage spreads to the internal network
L2	Low	It may cause damage to the external network, and restrict some services.
L1	Very Low	It may cause little damage to the external network and affect some services.

When analyzing the criteria for evaluating the network separation environment, the importance differs depending on the classification of the internal network and the external network. As such, protection of the internal network environment is a crucial criterion for network separation [8].

2.3. Network Separation Implementation Methods

The implementation method of network separation can be roughly classified into two methods. The first one involves controlling the network from the network's perspective. When blocking the blocked network based on the criterion of the supervisory control and data acquisition (SCADA), it must be designed with an air gap between the control network and the Enterprise LAN [9]. Unidirectional gateways called data-diodes are commonly used. This is a method in which a data exchange system is placed between blocked networks via non-TCP communication to allow files to be exchanged in a limited manner. The second method defines network separation from the user's perspective. The user uses two PCs. It consists of an internal network PC that is not connected to the Internet and an external network. In network separation, in order to utilize information on the Internet, both internal and external PCs must be used to transfer information. In this process, work efficiency is reduced.

A PC that is connected to the Internet, whether physical PC or logical PC, can be implemented according to the environment; the detailed classification is as shown in Figure 2.

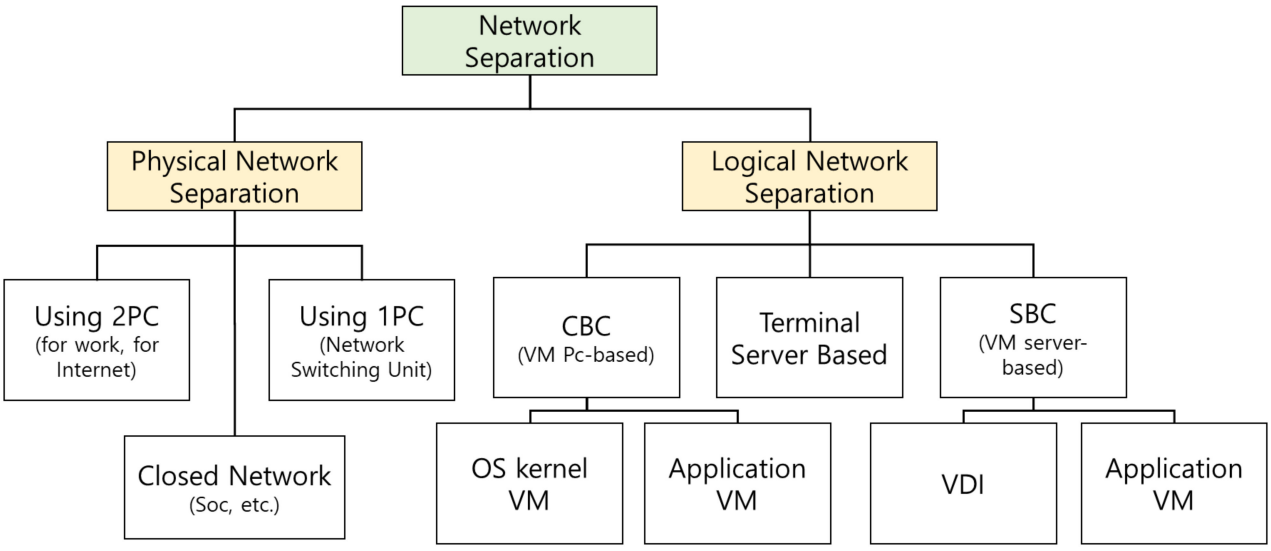


Figure 2. Diagram of network separation classification.

2.4. Research Trend Analysis on Network Separation

Hwang Sung-kyu's study [10] pointed out the problem of service failure caused by malware as the possibility of internal information leakage increases when a work PC is directly connected to an external network and access from the external network to the server is free. To solve the problem, he studied the network separation technique using network virtualization by classifying the security system area and the manganese data-linked system area.

Kim Il-Yong's study [11] pointed out the problem of damage caused by cyberattacks as it is connected to information and communication services, unlike in the past, which was only done in closed internal networks in existing industrial control systems.

Ji Jung-Eun's study [12] noted that the rapid development of the Internet requires network separation technology to protect important information from cyberattacks, such as hacking and malware.

3. Proposed Network Separation Protocol

3.1. Requirements Analysis

In this chapter, we analyze the legal analysis of the network connection equipment applicable in Korea by analyzing the domestic law in Korea and propose a new network separation model applicable to the network connection equipment via major vulnerabilities that occur in the end client's web browser and its defense system.

3.1.1. Laws Related to Network Separation in Korea

In Korea, the definition of network separation that should be implemented by companies dealing with information and communication infrastructure, information and communication networks, and financial information is shown in Table 2 (some laws include alternatives).

Table 2. Korean laws and regulations related to network separation.

Law/Regulation	Legal Text
ACT ON THE PROTECTION OF INFORMATION AND COMMUNICATIONS INFRASTRUCTURE [13]	Article 5 (Establishment of Measures to Protect Critical Information and Communications'Infrastructure) (1) The head of an organization that manages critical information and communications infrastructure (hereafter referred to as "management organization") shall formulate and implement management measures, including physical and technological measures, such as those for incident prevention, backup, and recovery to securely protect critical information and communications infrastructure and management information under his or her jurisdiction (hereafter referred to as "measures to protect critical information and communications infrastructure"), based on the outcomes of the analysis and evaluation of vulnerabilities under Article 9 (1) or (2).

Table 2. *Cont.*

Law/Regulation	Legal Text
PERSONAL INFORMATION PROTECTION ACT [14]	Article 29 (Duty of Safeguards) Every personal information controller shall take such technical, managerial, and physical measures as establishing an internal management plan and preserving access records, for example. that are necessary to ensure safety as prescribed by Presidential Decree so that the personal information may not be lost, stolen, divulged, forged, altered, or damaged.
	Article 30 (Establishment and Disclosure of Privacy Policy) (1) Every personal information controller shall establish a personal information processing policy including the following matters (hereafter referred to as "Privacy Policy"). In such cases, public institutions shall establish the Privacy Policy for the personal information files to be registered pursuant to Article 32:
ELECTRONIC FINANCIAL TRANSACTIONS ACT [15]	Article 21 (Duty to Ensure Safety) (1) A financial company or an electronic financial business entity and its or his or her subsidiary electronic financial business entity (hereafter referred to as "financial company, for example.") shall perform its or his or her duties of a good manager to ensure the safe processing of electronic financial transactions. (2) In order to ensure the safety and reliability of electronic financial transactions, a financial company for example. shall comply with the standards determined by the Financial Services Commission with respect to the information technology sector, such as human resources, facilities, electronic apparatus, and expenses for conducting electronic transmissions or processing, the electronic financial affairs and certification methods including the use of certificates under the Digital Signature Act.

Table 2. *Cont.*

Law/Regulation	Legal Text
ENFORCEMENT DECREE OF THE ACT ON PROMOTION OF INFORMATION AND COMMUNICATIONS NETWORK UTILIZATION AND INFORMATION PROTECTION, FOR EXAMPLE [16].	Article 15 (Protective Measures for Personal Information) (2) Each provider of information and communications services shall take the following measures to block illegal access to personal information pursuant to
	Article 28 (1) 2 of the Act: Provided, that a provider of information and communications services is obliged to take a measure under subparagraph 3, only if the number of users whose personal information has been stored and managed by the provider of information and communications services during three months immediately preceding the end of the previous year averages at least one million persons per day or the sales of information and communications services during the preceding year (referring to the preceding business year, if the service provider is a corporation) amount to at least ten billion won: 3. Blockade of external Internet networks to computers, for example. of persons accessing the personal information processing system while handling personal information;

In Korea, in accordance with the "Act on the Protection of Information of Information and Communications Infrastructure," organizations that manage crucial information and communication infrastructure are required to establish and implement physical and technical measures to prevent accidents in the infrastructure and prepare for future accidents. In addition, in accordance with the "Personal Information Protection Act," the representative who processes personal information is responsible for collecting internal management plans against the loss and theft of personal information and taking measures against leakage or forgery of records. In particular, the "Enforcement Decree of the Act on Promotion of Information and Communications Network Utilization and Information Protection, for example" defines that, in a system that handles personal information, the user's device accessing the system can be used only on the internal intranet where the external network is blocked.

As such, in Korea, heavy security is required for personal information, and network separation is mandatory for preventing access to external networks by systems that manage personal information. However, the current network connection system can be a problem in terms of work efficiency, and a more effective network connection system is required because work efficiency is directly related to the financial part.

3.1.2. Configuration of Network Connection System and Its Limitations

In a separation isolation environment, a service that can transfer data and files between the non-secure area and the secure area is required. A secure network transmission system is required to leverage external data necessary for work while maintaining heavy security of network separation. The data connection device commonly uses the encryption channel for all sections for secure data transmission in the network and supports features preventing the influx of malicious code via antivirus software when importing to the internal network and preventing the outflux of internal data via the approval of the administrator when exporting to the external network [17]. The streaming service provides a transmission feature via its non-TCP/IP-based protocol encryption/decryption process

for the separated internal and external network sections. It provides a connection service in accordance with security policies registered in the internal network relay server in real-time for data that need to be connected and guarantees the highest transmission speed with its high-speed data processing technology. Depending on the environment, it provides a method of real-time service connection (stream connection) in national defense and public institutions that requires a robust security configuration via server-to-server file connection. Instead of allowing connections between the internal and external networks, procedures are required to ensure the safety of the transmitted files [18]: first, the transmission of files only approved by the administrator (including post-approval); second, identification of known malicious codes via antivirus software; third, behavioral anomaly detection using SandBox-based tools.

By following the above procedures, the security level can be ensured via network separation. However, when leveraging information from a site, more complex procedures may be required to determine whether malware is present for access to other sites, such as hyperlinks, which may require more performance time.

3.1.3. Definition of Major Vulnerabilities and Defense System That Occur in Web Browsing

This section analyzes OWASP10 and MITRE Top 25 used in web vulnerability management systems to analyze security vulnerability items that may occur on internal networks in network-isolated environments and sets them as safety criteria for the proposed network-isolated environment.

The OWASP is an open-source web application security project that mainly studies information disclosure related to the web, malicious files and scripts, and security vulnerabilities. Vulnerabilities that have a high frequency of occurrence or a significant impact on security among web application vulnerabilities are announced in the OWASP TOP 10.

MITRE, which aims to strengthen security via research on security vulnerabilities, has published the results of its analysis of vulnerabilities called MITRE releases 2020 CWE Top 25. MITRE announced its list of the top 25 most likely problems, including software problems (errors, bugs and potential attack vectors), system hijacking, data breaches (and theft of sensitive data), denial-of-service (DoS) attacks, system crashes, and arbitrary code execution [19].

We selected four (shown in Table 3 below) that target the endpoint among the eleven that have a high risk and frequency as selected by the OWASP Top10 and MITRE Top 25.

Table 3. Examples of high-risk web vulnerabilities.

Rank	Vulnerability	Detail Description	Exploit	OWASP	MITRE
1	CVE-2014-6271	Execution of arbitrary code through crafted environment	(){:;};ping-c1-p cb18cb3f7bca4441a595fcc1e 240deb0 attacker-machine.com	A9	11
2	CVE-2014-6278	Execution of arbitrary code through crafted environment	() {:;}; /bin/sleep 20 | /sbin/sleep 20 | /usr/bin/sleep 20		
3	CVE-2014-6277	Execution of arbitrary code through crafted environment	'f() { x() { _;}; x() { _;} <		
4	CVE-2017-5638	Remote execution of arbitrary commands through crafted content-type HTTP header	payload += "(#cmd = '%s')." % cmd try:headers = {'User-Agent':'Mozilla/5.0','Content-Typ e': payload}		3

Vulnerabilities in web browsers are discovered every year. In this paper, we analyze four weaknesses: CVE-2014-6271, CVE-2014-6278, CVE-2014-6277, and CVE-2017-5638, for which we analyze the vulnerabilities of OWASP and MITRE.

CVE-2014-6271 is a vulnerability in which the validation of the input value is not performed normally when a command registered abnormally is executed in the process of registering environment variables by Bash, the Unix shell, and the script running on the DHCP client on Apache HTTP server is executed outside of privileges of the bash [20]. CVE-2014-6278 is a vulnerability in which code is executed by an environment variable caused by a vulnerability that occurs by bypassing the patch even after the patch for CVE-2014-6271, described above [21]. CVE-2014-6277 is a vulnerability in which GNU bash cannot analyze the definition of environment variables, causing arbitrary code execution or denial-of-service by attackers [22]. Finally, for CVE-2017-5638 using object graph navigation language (OGNL) expression, Apache Struts can use remote code in most processes, and it takes advantage of this to remotely execute arbitrary code [23]. The analyzed vulnerabilities are those that occur in the shell that is currently being used and are commonly caused by certain environment variables that lead to attacks. The above vulnerabilities are similar to those selected as vulnerabilities identified in OWASP and MITRE CWE Top 25 [24,25]. Using components with known vulnerabilities in A9 of the OWASP Top 10 is an item for a vulnerability that exploits the code of known vulnerabilities and belongs to A9 [26,27].

Of the MITRE CWE Top 25, 3 and 11 are improper input validation and integer overflow or wraparound, respectively, and improper input validation refers to a vulnerability that occurs in the process of receiving input data by failing to verify that the data have the properties required to process it safely and correctly. CVE-2017-5638 belongs to this in that it allows code to be executed remotely without ensuring the safety of the received data. Next is an integer overflow or wraparound, which is a vulnerability that occurs when an integer value increases to a very large value when storing its content. This is a vulnerability because the overflowed value can be a break or a negative break, and it occurs in a value that deviates from the value recorded in the environment variable in the bash shell. The vulnerabilities that occurred in CVE-2014-6271, CVE-2014-6278, and CVE-2014-6277 belong to an integer overflow or wraparound.

3.2. Network Flow Design

The proposed web-browsing application model can be applied to a PC-based network separation model and applies to both physical and logical network separation. However, its structure requires a data exchange system. Since there is a CC criterion for the safety criterion of data exchange, we proceeded with the design assuming that data exchange is safe. Figure 3 shows the overall flow of the proposed model. Non-TCP encryption communication is used to transfer files between the internal network and the external network. For files transferred from the outside to the inside, the signature of malicious code is searched for through anti-virus for prevention work, and malicious code is searched for through the SandBox. It judges the behavior and verifies the malware through pattern matching using YARA framework, a tool to identify and classify types of malware.

3.2.1. Diagram of the Network Flow Sequence of the Proposed Model

When using the internal network system in the internal network, it generally performs the web service in the same way as the current network separation policy, as shown in Figure 4. To translate the URL entered in the browser into an IP address, it makes a DNS query to the private DNS in the internal network, and performs web service to the IP address that received the query response.

However, using the external network system (Internet) from the internal network is not possible due to the current network separation policy, therefore, to import files into the internal network, it is necessary to access the external network PC, access the Internet website, make the necessary information into a file, and perform the file transmission process. This process wastes time and causes inefficiency. The proposed model defines the

network flow diagram as below to display the files in the internal browser by transferring unsearched URLs to the external and transforming them into PDF, for example. When querying the private DNS from the internal network. When configured this way, additional features, such as hyperlinks to the website can be removed. This removes the process of importing internally and improves efficiency by preventing the execution of malicious code that exists in the data and simultaneously creates a website as a separate file.

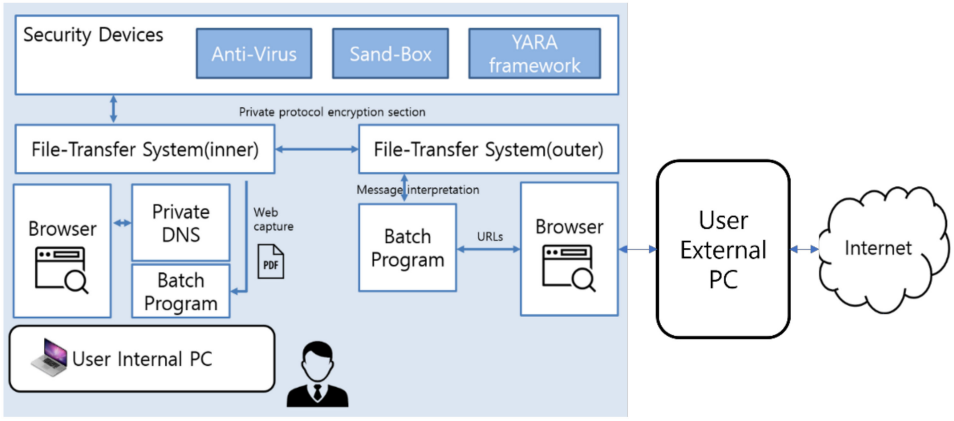


Figure 3. System model of proposed web scraping platform.

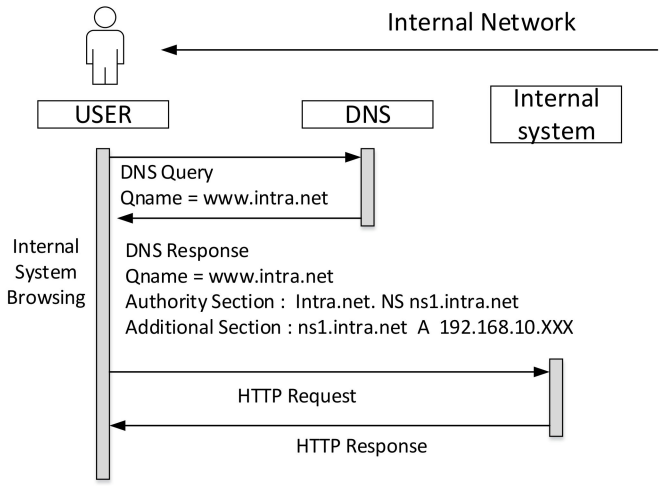


Figure 4. Internal system request sequence diagram.

The agent may be designed and operated by the proxy method and driven by a PC in the external network. Figure 5 shows the network flow of the proposed model where the agent in the provided model is configured to operate in the following order. The reason not to use the protocol used for web services is to neutralize vulnerabilities of existing web services.

a. Wait for the URL forwarding message.
b. Read the message when it is delivered.
c. Render the page using the Headless browser library.
d. Download the web according to the requested renderType.
e. Move the file to the sync folder.
f. Move the sync folder internally via data exchange.
g. Run the received file (for PDF files, open it with a browser).

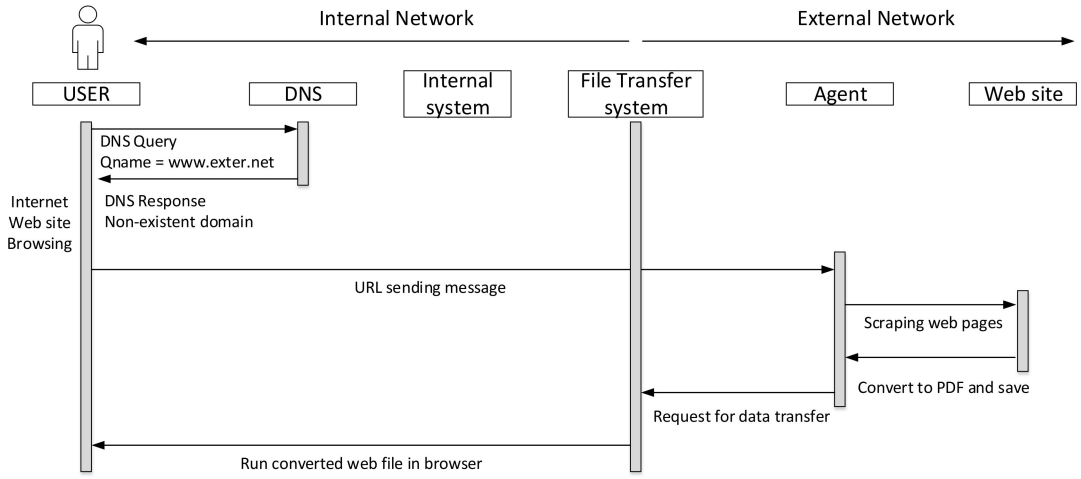


Figure 5. Internet web request sequence diagram.

3.2.2. Designation of DNS

In a network separation environment, the internal network is unable to access the external DNS server, and, therefore, unable to perform a DNS query. Therefore, the DNS server located on the internal network must return the IP result of a certain DNS without having to communicate with the external, and we designed the protocol to execute this as shown in Figure 6.

Explanation of terms
INPUT: URL entered in the browser to query DNS and transform to IP
e.g.,) <https://www.intra.net>
OUTPUT: The result value of the received value (IP, for example.) after querying DNS for the requested URL.
Internal network web service: www.intra.net(IP addr: 192.168.10.XXX)
Public web service: www.exter.net(IP addr: 220.111.11.XXX)

Private DNS configuration algorithm
INPUT: User input URL
OUTPUT: IP(or Error message)
1. if DNS Recursive Name Server[List] ≠ DNS query
2. Forward to external network agent (via. File Transfer system)
3. Query a URL in External DNS
4. if External DNS Recursive Name Server[List] ≠ DNS query
5. return(Error message: No records, Nonexistent domain)
6. else
7. return(Capture Web of URL)
8. else
9. return(IP of requested URL)

Figure 6. Private DNS configuration algorithm.

3.2.3. Definition of URL Forwarding Message Protocol

It is used to deliver the required information from the internal to the external by defining a protocol that delivers the URL to the external network agent, and delivers the scraped file as a response. The URL messaging protocol is defined as JSON type as follows. The property of required information is the information required for scraping from the

Internet network, defines the URL and return type and designates the scraping method. Finally, it takes a procedure to authenticate the minimum number of senders using authKey.

4. Implementation and Test Analysis

4.1. Web-Browsing Application Implementation

In this section, we perform an analysis to verify the effectiveness of the proposed model. Figure 7 shows a comparison of the *New York Times* between browser and scraped material. It was rendered slightly differently because the responsive web design was applied depending on the viewport applied to the html, but it can be seen that the content and quality of the two are mostly the same.

Figure 7. (**Up**) browser web, (**Down**) scraping web.

The agent device that performs web scraping using the resources of the Internet PC is configured as shown in Figure 8, which also shows the source program for configuring the actual agent. The code implements the web-scraping process using the resources of the external network PC. Invokes the headless browser package to scrape the web in the background. Because files must be generated, transferred, and deleted, a programming language is required to control OS files. It was developed using a programming language that meets the requirements for the creation of experimental programs.

4.2. Analysis of Malicious Code Safety

As a result of analyzing the form of HEXA code, it can be determined that obj, which is the extension of the stored records, brought the hyperlink of the scraped web page as it is. Figure 9 is the analysis result of HEXA code form.

As a result of converting to Hex code using an editor that can analyze PDF and analyzing it as in Figure 10, it can be seen that the hyperlink implemented in the original web page was completely scraped in its intact form. In this way, the scraped material can import the hyperlinks contained in the original web page, but a safety analysis of those hyperlinks must be performed. In this paper, a drive-by download test was performed to verify the security of the addresses contained in the hyperlinks. Drive-by download refers to an attack that is performed when a user visits a specific website by installing malicious code on the user's device without the user's knowledge. When accessing a site that implements drive-by download using the <frame> tag to perform a drive-by download attack, the browser calls the URL inside the frame tag. This URL causes the download to execute the downloaded file. Downloading a specific file begins without pressing any download or save button from the test web page via running the downloaded file.

```
#include <file.au3>

$fileNM = ""
$urlFileNM = ""
Exec()
moveFile()

Func Exec()
    Local $hSearch = FileFindFirstFile("*.adr")
    If $hSearch = -1 Then
        Return False
    EndIf

    Local $urlFileName = ""
    $urlFileName = FileFindNextFile($hSearch)
    $urlFileNM = $urlFileName
    Local $urlFileopen = FileOpen($urlFileName, 0)
    Local $sFileRead = FileRead($urlFileopen)
    FileClose($urlFileopen)
    ;Run application
    Run("cmd.exe")
    ;Wait for CMD to be opened
    WinWaitActive("Administrator: C:\Windows\system32\cmd.exe", "", 1)
    Send('cd O:\사후결재받은함' & "{ENTER}")
    $fileNo = Random(0, 100, 1)
    $fileNM = 'screen' & $fileNo & '.pdf'
    Send('phantomjs rasterize.js ' & $sFileRead & ' screen' & $fileNo & '.pdf' & "{ENTER}")
    sleep(9000)
EndFunc

Func moveFile()
    FileMove($fileNM, 'O:\보낼파일함\' & $fileNM & '.jpg', $FC_OVERWRITE) ; $FC_OVERWRITE (1) = overwrite existing files.
    FileDelete($urlFileNM)
EndFunc
```

Figure 8. Web scraping agent device.

85 0 obj
<</Type/Annot
/Subtype/Link
/F 4
/Border [000]
/Rect [28.5 4926 931.5 4982.25]
/A <</Type/Action
/S/URI
/URI (https://www.nytimes.com/live/2021/02/10/us/impeachment-trial/?action =
click&module = Spotlight&pgtype = Homepage)>>>>
endobj

Figure 9. Result of PDF to HEX conversion.

Figure 11 shows the contents that traced the process of being loaded using the actual browser network development tool, and the lower row part shows the code that drive-by download performs using the frame tag.

Figure 12 shows that when accessing the drive-by download site, the anti-virus judges malicious behavior and immediately blocks its execution. If it is an unknown malicious code, the payload will succeed and the hacker (attacker) will be able to execute it because the anti-virus software makes a judgment by looking at the signature of known malicious codes.

When opening a file saved as PDF with a browser after scraping, it can be seen that drive-by download did not work, and the JS and JavaScript code that can also work with frame tag and PDF was removed, as shown in Figure 13.

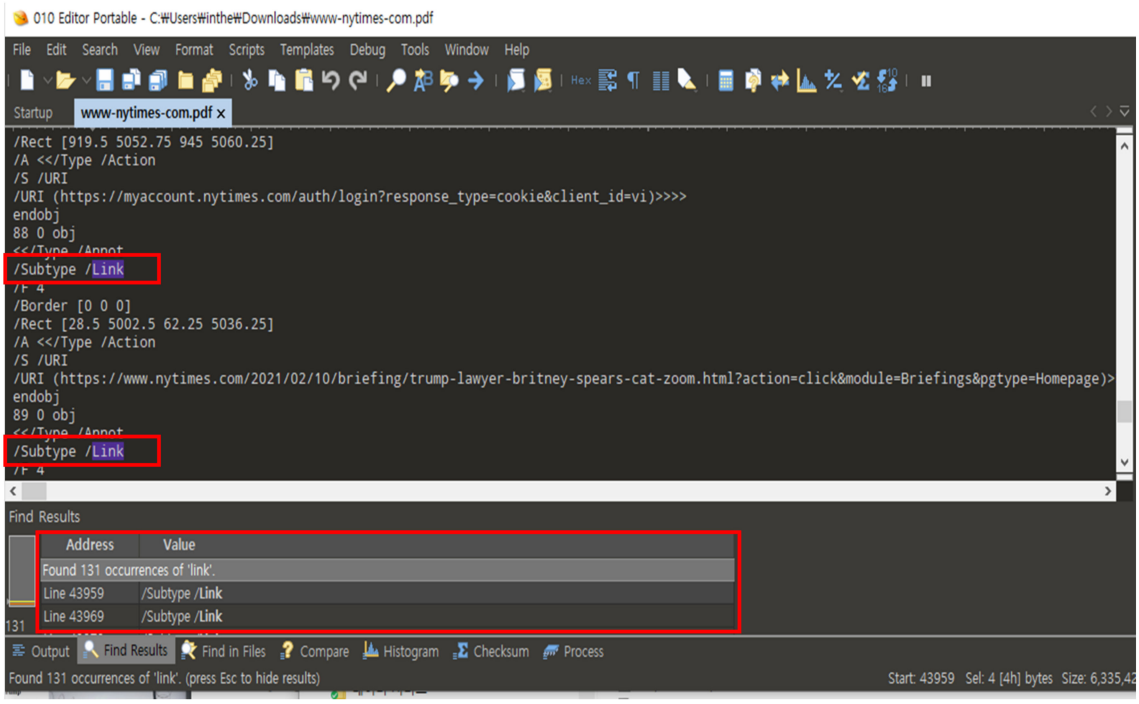
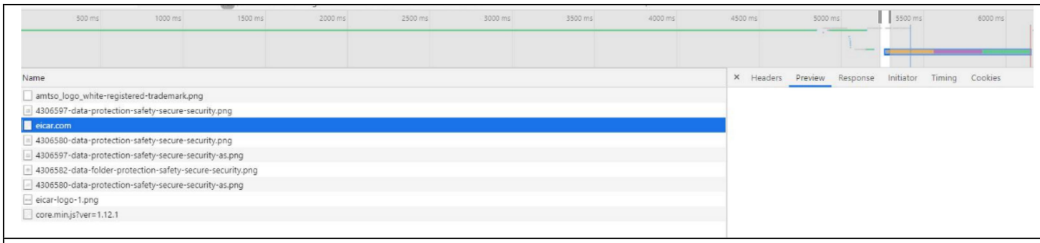


Figure 10. Convert PDF to Hex code for analysis.



<div class = “elementor-element elementor-element-6eaa952c elementor-widget elementor-widget-html” data-id = “6eaa952c” data-element_type = “widget” data-widget_type = “html.default”>

<iframe loading = “lazy” src = “https://amtso.eicar.org/eicar.com” width = “1” height = “1” frameborder = “0”></iframe>

Figure 11. Convert to PDF by scraping.

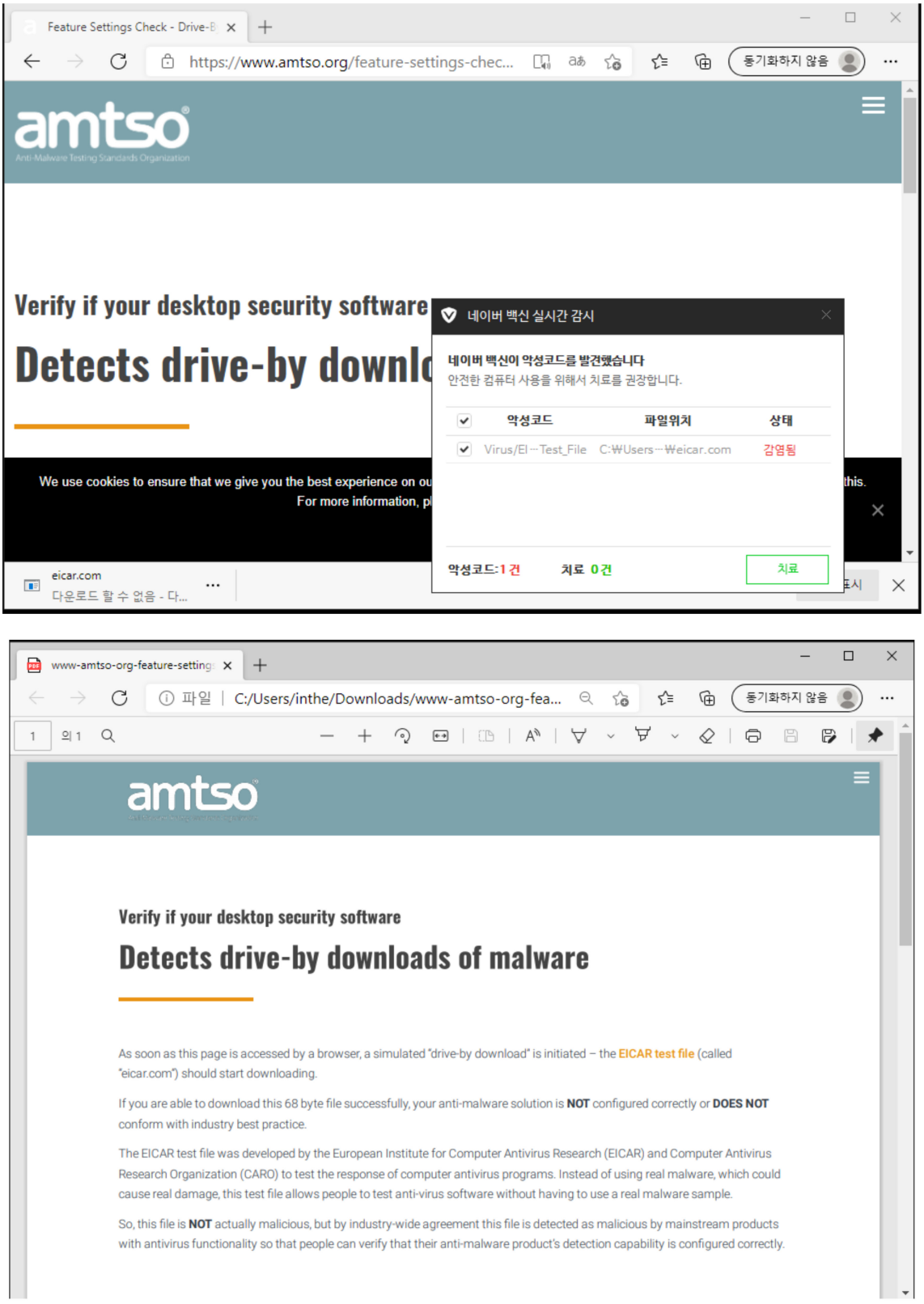


Figure 12. (**Up**) 'Drive-by Download' test web (**Down**) convert to PDF by scraping.

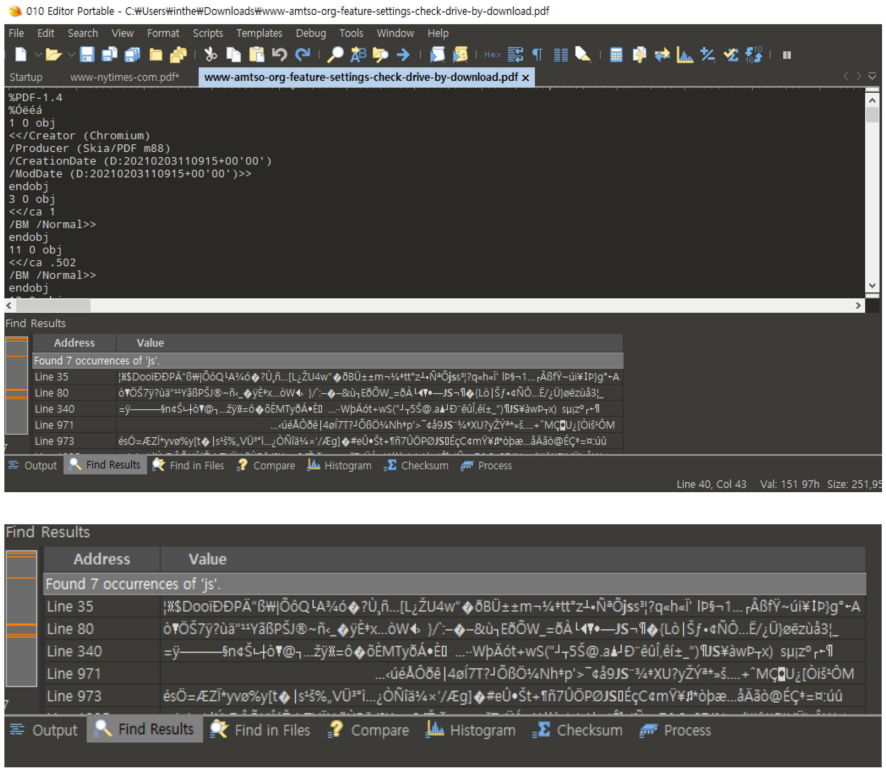


Figure 13. (**Up**) image information of pdf rendering materials (**Down**) JavaScript control information.

The URL is not called in the frame because the frame is captured in the file converted to PDF. It will not be possible to download the file because it is not behavior that occurs in the browser, even when it is called.

It can be seen from Figure 12 that many tags that operate in the browser have been transformed into still images or texts, and the working code remains like a link tag. In Figure 13, the pdf-rendered data of the site is converted into hexacode, and it can be seen that the pdf image is converted into text form in a. Additionally, in b of Figure 12, you can see that the JavaScript that can execute the logic has been.

Table 4 shows a comparison of the automated web scraping methods we propose and the manually operated web scraping method. The rendering rate for the web itself shows the same results in both ways, but the anti-vaccine, SandBox inspection time shows that there is no time required. This takes time to determine whether malware exists on the web in the case of manual methods, but the proposed method eliminates the malware contained on the web, converts it into pdf, and transmits it, so no separate malware detection is required. To compare the two techniques, we used a drive-by-download test to determine the time of manual malware detection, and performed the code produced to average the time required to run.

Table 4. Comparison of scraping methods.

	Web Page Rendering	**Anti-Vaccine, SandBox Inspection Time**	**Transmission Time**
Manual capture method	3 s	more than 60 s	10 s
Automated scraping method	same as above	Not applicable	same as above

5. Conclusions and Future Plans

Many companies and organizations are applying or planning to separate networks in response to increasing cyberattacks and threats. In particular, due to COVID-19, research on models for building a non-face-to-face smart work environment in a network separation environment is also being started. It is necessary to consider the efficient aspect of network separation technology that is in line with ICT technology to develop into a super-connected society.

A search of the Korean Academic Journal Citation Index (KCI) for the last five years using the keyword "network separation" produces fifty-three papers. Most of the research on network separation is focused on strengthening security and blocking malicious code. The purpose of this paper is to propose a model that allows safe web browsing in a network separation environment. This paper proposed a way to work more efficiently in a network separation environment, and confirmed that the captured web transmitted by non-TCP data exchange cannot operate with malicious code. Additionally, since most PDFs are opened in a web browser and scraped in the form of working hyperlinks, it is proven that similar web browsing can be experienced in a static state.

The proposed technique is a method of transferring information from the site to the pdf with the hyperlink removed preferentially when accessing an unlisted site for network separation. Commonly used techniques take more time to run a virus presence check on each hyperlink in a site, while the proposed method takes less time to get only the content of the site except for the hyperlink.

In the future, the html browsing language will be defined and implemented in raw form to reduce the time spent converting information from the site to pdf. This approach is expected to reduce the time required by reducing the process of rendering html information on the site in pdf and allowing access to information on the site immediately.

Author Contributions: Conceptualization, W.-C.J.; methodology, W.-C.J. and J.K.; software, W.-C.J.; validation, W.-C.J., J.K. and N.P.; formal analysis, J.K.; in-vestigation, W.-C.J.; resources, W.-C.J. and N.P.; data curation, J.K.; writing—original draft preparation, W.-C.J.; writing—review and editing, J.K. and N.P.; visualization, W.-C.J.; supervision, N.P.; project administration, N.P.; funding acquisition, N.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Korea Foundation for the Advancement of Science and Creativity (KOFAC) grant funded by the Korea government (MOE).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kim, C.S. Policy Study on Elimination of Unnecessary Traffic in Network Detachment Environment. Master's Thesis, Korea University, Seoul, Korea, 2019.
2. Pavithra, P.; Hartwig, H.; David, L. Network Structure and Spatial Separation. *Environ. Plan. B Urban Anal. City Sci.* **2012**, *39*, 137–154.
3. Henner, V.E. A network separation scheme for emergency control. *Int. J. Electr. Power Energy Syst.* **1980**, *2*, 109–114. [CrossRef]
4. Lee, H.J.; Cho, D.I.; Kou, K.S. A Study of Unidirectional Data Transmission System Security Model for Secure Data transmission in Separated Network. *Asia Pac. J. Multimed. Serv. Converg. Art Humanit. Sociol.* **2015**, *5*, 539–549. [CrossRef]
5. Rajendra, P.; Harsha, D.; Chirag, M. Designing an efficient security framework for detecting intrusions in virtual network of cloud computing. *Comput. Secur.* **2019**, *85*, 402–422.
6. Lee, D.H.; Kim, H.K. Real-time security Monitoring assessment model for cybersecurity vulnerabilities in network separation situations. *J. Inf. Secur.* **2021**, *21*, 45–53.
7. ISO 27001 Network Segmentation Overview. Available online: https://iso27001guide.com/iso-27001-network-segmentation-overview-iso27001-guide-iso27001-guide.html (accessed on 26 May 2021).
8. Jung, W.C.; Park, N. A Safe Web in Network Separation Environment. *J. Comput. Theor. Nanosci.* **2020**, *17*, 3243–3249. [CrossRef]

9. Jung, W.C.; Park, J.; Park, N. Safe Web Using Scrapable Headless Browser in Network Separation Environment. *J. Korea Soc. Comput. Inf.* **2019**, *24*, 77–85.
10. Hwang, S.K. Network separation construction method using network virtualization. *J. Korea Inst. Inf. Commun. Eng.* **2020**, *24*, 1071–1076.
11. Kim, I.Y.; Lim, H.T.; Ji, D.B.; Park, J.P. A Efficient Network Security Management Model in Industrial Control System Environments. *Korea Acad. Ind. Coop. Soc.* **2018**, *19*, 664–673.
12. Jee, J.E.; Lee, S.G.; Lee, S.R.; Bae, B.C.; Shin, Y.T. A Logical Network Partition Scheme for Cyber Hacking and Terror Attacks. *Korean Inst. Inf. Sci. Eng.* **2012**, *39*, 95–101.
13. Act on the Protection of Information and Communications Infrastructure. Available online: https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=136754&viewCls=engLsInfoR&urlMode=engLsInfoR&chrClsCd=010202&lsId=009182#0000 (accessed on 19 August 2021).
14. Personal Information Protection Act. Available online: https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=53044&type=lawname&key=personal+information (accessed on 19 August 2021).
15. Electronic Financial Trans-Actions Act. Available online: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=44455&lang=ENG (accessed on 19 August 2021).
16. Enforcement Decree of the Act on Promotion of Information and Communications Network Utilization and Information Protection, Etc. Available online: https://elaw.klri.re.kr/kor_mobile/viewer.do?hseq=42587&type=part&key=43 (accessed on 19 August 2021).
17. Jung, W.C. Technical Scheme for Network Separation: Focusing on the Web Data Transmission. Master's Thesis, Jeju National University, Jeju, Korea, 2020.
18. Ali, M.; Pierluigi, N.; Alberto, L.S.V.; Jan, M.R. Optimized Design of a Human Intranet Network. In Proceedings of the 2017 54th ACM/EDAC/IEEE Design Automation Conference (DAC), Austin, TX, USA, 18–22 June 2017.
19. Cho, B.J.; Yun, J.H.; Lee, K.H. Study of effectiveness for the network separation policy of financial companies. *J. Korea Inst. Inf. Secur. Cryptol.* **2015**, *25*, 181–195.
20. CVE-2014-6271. Available online: https://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2014-6271 (accessed on 26 May 2021).
21. CVE-2014-6277. Available online: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-6277 (accessed on 26 May 2021).
22. CVE-2014-6278. Available online: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2014-6278 (accessed on 26 May 2021).
23. CVE-2017-5638. Available online: https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2017-5638 (accessed on 26 May 2021).
24. Arpana, K.; Vivek, K. Competing secure text encryption in intranet using elliptic curve cryptography. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 631–641.
25. A9:2017-Using Components with Known Vulnerabilities. Available online: https://owasp.org/www-project-top-ten/2017/A9_2017-Using_Components_with_Known_Vulnerabilities (accessed on 26 May 2021).
26. CWE-20: Improper Input Validation. Available online: https://cwe.mitre.org/data/definitions/20.html (accessed on 26 May 2021).
27. CWE-190: Integer Overflow or Wraparound. Available online: https://cwe.mitre.org/data/definitions/190.html (accessed on 26 May 2021).