TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Aaditya Parashar 177778IVSB

A WEB-BASED PLATFORM FOR LEARNING AND PRACTICING CYBERSECURITY

Bachelor's Thesis

Supervisor: Kaido Kikkas
PhD

Tallinn 2020

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Aaditya Parashar 177778IVSB

VEEBIPÕHINE PLATVORM KÜBERTURBE ÕPPIMISEKS JA HARJUTAMISEKS

bakalaureusetöö

Juhendaja: Kaido Kikkas
PhD

Tallinn 2020

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Aaditya Parashar

18.05.2020

Abstract

This work aims to analyse and develop a web-based platform for learning and practicing cybersecurity. This work gives an understanding of why is there a need for such a platform and how relevant the subject of cybersecurity is in the society currently.

The idea is not to replace the university or school level education or offer accredited certifications; the idea is to have a web platform which everyone, including schools and universities, can use to enhance their education quality in the field of cybersecurity.

Furthermore, the work analyses existing solutions and learning methods used in e-learning, and based on those learnings the solution development was analysed which included analysing the language to use and the framework to use for the solution, as well as what software will be in use in the infrastructure components.

On completing the analysis, the development took place for a web platform prototype. This work includes the implementation description of the prototype.

This thesis is written in English and is 48 pages long, including 7 chapters, and 12 figures.

Annotatsioon

Veebipõhine platvorm küberturbe õppimiseks ja harjutamiseks

Käesoleva töö eesmärgiks on veebipõhise küberturbe õppe- ja arendusplatvormi analüüs ja arendus. Uuritakse, miks sellist platvormi on vaja ning milline on hetkel küberturbe roll ühiskonnas.

Eesmärk ei ole asendada kooli- või ülikooliharidust ega täiendõpet, vaid pakkuda platvormi kõigile huvilistele (sh koolid ja ülikoolid) küberturbealase hariduse täiendamiseks. Lisaks on siin analüüsitud olemasolevaid lahendusi ja e-õppe meetodeid ning selle põhjal pakutud välja ideid vastava taristu arendamiseks (programmeerimiskeel, tarkvararaamistik, 5ng lahendused).

Analüüsi tulemusena arendati veebiplatvormi prototüüp, mille rakendamist on töös kirjeldatud.

Lõputöö on kirjutatud 5nglese keeles ning sisaldab teksti 48 leheküljel, 7 peatükki ja 12 joonist.

List of abbreviations and terms

API	Application Programming Interface
BYOD	Bring Your Own Device
CAGR	Compounded Annual Growth Rate
DB	Database
E-learning	Electronic Learning
HTML	Hypertext Markup Language
IP	Internet Protocol
IT	Information Technology
MIT	Massachusetts Institute of Technology
OWASP	Open Web Application Security Project
PVE	Proxmox Virtual Environment
RDBMS	Relational Database Management System
REST	Representational State Transfer
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SQL	Structured Query Language
SWIFT	Society for Worldwide Interbank Financial Telecommunication
USA	United States of America
VM	Virtual Machine
VPN	Virtual Private Network

Table of contents

List of figures

1 Introduction

This work deals with the analysis and development of a web-based platform for cybersecurity learning and practising. This work includes analysing the current solutions, analysing the teaching methodologies for digital learning, and developing a proof-of-concept of a web-platform for the solution.

The target for the solution is to show how such an intended solution would benefit and solve the problem of not having a one-platform solution available to the public. The solution includes analysis and development of the infrastructure elements, which include:

- Webserver
- Application Server (Core Server)
- Database Server
- Hypervisor
- Proxy Solution

While learning can happen by using some technical resources available online and setting up ones practice environment; it is not an ideal way, especially for new students who mostly have no experience in doing such a set-up. Additionally, the resources available on the internet might not be comprehensive and could potentially include incorrect or outdated information. The idea here is not to replace the university or school level education or offer accredited certifications; the idea is to have a web platform which everyone, including schools and universities, can use to enhance their education quality in the field of cybersecurity.

The thesis contains the following chapters:

Chapter 2 deals with discussing the problem this work aims to solve; that is done by looking at the relevance of the problem in the real world by describing some case-studies and summarising them.

Chapter 3 goes into the analysis of the current solutions to the problem and a comparison among them. Additionally, this chapter includes a brief market analysis for potential markets which such an intended solution would be highly beneficial.

Chapter 4 goes into analysis of various teaching methodologies which are used for digital learning; this chapter also includes an analysis on a survey created by the author to get a first-hand account of the effectiveness of the solutions analysed in this chapter.

Chapter 5 goes into the analysis of how the ideal solution should be developed, analysing topics like the language and framework, as well as the software for the infrastructure elements like web servers and database servers, which should be used for the solution.

Chapter 6 describes the actual development of the solution, how the different use cases of the platform are handled on the server-side, and how the deployment for the practical labs happens.

2 The problem and its relevance

The problem this thesis is aiming to solve is that there is a lack of platforms available online where students and enthusiasts can register independently and start learning and practising skills in the cybersecurity domain.

This chapter looks at the relevance of this problem in today's society by looking into certain case studies of cybersecurity failures at the corporate, and government. Additionally, before diving into the case studies of lapses of cybersecurity, perhaps its best to understand what it is and why it is needed.

2.1 What is cybersecurity and why it is needed?

Cybersecurity is a term which we have seen being used more and more in the recent years, hence let us start by defining it. The technical definition is *"Precautions taken to guard against crime that involves the Internet, especially unauthorised access to computer systems and data connected to the Internet"* [1]. Although, that definition could be improved, as it defines cybersecurity only as some actions or precautions which one takes, which in-part is correct indeed. In addition to the actions one takes to guard against crimes, it should include a thought process, almost a sub-culture of development that has this view of cybersecurity ever-present.

Now that we have established what the term cybersecurity means, perhaps we should also establish why it is necessary. The internet has evolved since it was made public in the 1990s, evolved in the sense of making things possible to be done digitally which until then could not be done online, things like shopping, communicating using e-mail, and eventually instant messaging. In the case of Estonia, even voting became something which can be done entirely digitally. Then in addition to just services, there were whole new industries created who provided services only digitally; services such as social networking with platforms like Myspace, Orkut, and eventually Facebook; search engines like Bing, Yahoo and Google; these are just some of the examples, but what all this meant was there were flocks of people going digital every day.

The common denominator for all these services, companies, or governments, even though they provide very different services is what they all rely upon – data. The data includes things such as names, photos, identity numbers, e-mail addresses, bank accounts, medical files, and more of the customers or citizens who use the services. Just as in the physical world, some groups or individuals have malicious intent to harm business, governments, or individuals for monetary or other reasons, they also exist in the digital world. It is because of such people or groups that the demand for cybersecurity has been increasing, as more people use the internet, the more data they store, the more the companies and governments need to protect their systems from unauthorised access.

2.2 The relevance of the problem in society

Now there two things established about cybersecurity – what it is and why it is becoming so popular in the recent time. Based on that, everyone would assume that it is something taken very seriously by every company and many individuals; however, the reality is far from it.

For instance, in the last decade alone, there have been significant security breaches, like the eBay hack in May 2014, exposing names addresses and more for its 145 million users. Alternatively, the Marriot International breach which started in 2014 and was announced in November 2018 where the attackers had stolen data on approximately 500 million customers. [2]

What these attacks show is that companies and organisations are not implementing cybersecurity practices properly. Below are some case-studies to explain how severe the lack of cybersecurity implementation and knowledge is currently.

2.2.1 Case study of a company

Taking the example of a company which is relatively a new player in an already crowded online market for video conferencing services; the other solutions which have been existing for many years, owned by companies which are very well established players in the technology industry. To differentiate themselves, this new company adds more features which make their service more interesting to use and more straightforward, which is very helpful and useful for many people. However, since it was a small platform and when it launched, it did not get a large number of users on it, and it never got attacked to

the extent of the other large platforms, this allowed to hide the platform's vulnerabilities for a while.

Then a crisis happened, and everyone needed a comfortable video conferencing platform for their work, for school, or just for communicating with their families. This new platform was suddenly thrust into the spotlight, and was getting more users than ever before; in 5 months, this platform went from 10 million users to over 300 million daily users ([3][7]). This growth was very good for business; however, with this increase in users, it also meant an increase in attacks against the platform.

This increase in attacks was a severe cause for concern as to how the development was done on the platform allowed for severe vulnerabilities to exist. Both the website and the applications had vulnerabilities, such as, how this platform had significant vulnerabilities in its installer, in its application on macOS, and even how it was possible to bypass the e-mail verification process [4].

If exploited, these vulnerabilities could be disastrous; the interesting learning from this was that all the vulnerabilities could have been avoided if security was given a high importance in the development life-cycle.

One would say it is unlikely that this situation could happen in 2020, especially in light of the increased focus on cybersecurity after attacks in the last decade. However, this example is in-fact true, and a platform which has found extensive use during the SARS-CoV-2 emergency in 2020, it is named – zoom. ([5][35])

2.2.2 Case study of a government

The above case-study showed poor security practices by a company affecting its users by making the calls and user data vulnerable. However, that did not have any direct enormous financial impact – which is what people seem to care about when talking about cyber threats. That enormous financial impact is what happened in February 2016, when a group of cybercriminals broke into the Bangladesh central bank and initiated 35 transactions amounting to approximately $1 billion [10].

What happened there was that the cybercriminals infected one of the bank systems by sending a malicious e-mail to an employee who opened that mail and compromised the bank systems. That in itself is bad, but what made it worse is that on the same network

the bank had its Society for Worldwide Interbank Financial Telecommunication terminal, colloquially called a SWIFT terminal.

SWIFT is an international network that enables financial institutions to send transaction orders in a secure, standardised, reliable, and trusted way. What that means is since SWIFT is a trusted system when the bank receives a transaction order it would assume the order to be valid and authorised and do not do any double-checking. What this meant was, the cybercriminals could get access to the Bangladesh central banks SWIFT terminal and could log in using the employee credentials they gathered from the infected system. They then proceeded by issuing 35 transaction orders to the New York Federal Reserve to transfer funds out of the Bangladesh central banks account there to international bank accounts belonging to the cybercriminals. The 35 orders totalled to US$951 million [10]; fortunately, 30 of the orders were flagged for manual review due to some words present in the orders were blacklisted in the USA. Additionally, another order was returned due to a spelling error. However, the remaining four orders went through and cost the bank US$81 million ([8][9]).

This entire hack could have been avoided by implementing proper cybersecurity practices, one of the easiest things to have prevented this was isolating the SWIFT terminal, as recommended by the SWIFT organisation itself to prevent exactly this type of attack. ([11][12])

2.2.3 Summarising the problem

As a summary of the problem, what was established is that although it is becoming more and more critical for companies and governments to protect their cyber interests, there are certainly deficiencies in that department. The way a company decides to implement cybersecurity practices is dependent on the company; one of the ways is by enhancing the knowledge of everyone developing and administrating IT systems. One way that can be achieved is by using online learning platforms where users can learn and experience a range of cybersecurity-related topics, which in reality would hopefully not happen.

Additionally, it is highly beneficial to start focusing on such training and practical experiences from the school level itself, as that is where students are learning IT concepts and at the same time learning cybersecurity concepts imbibes the security mindset from the start.

3 Analysis into existing solutions and markets

This chapter analyses the existing solutions for the problem, and it describes what those solutions offer and what are their limitations. Additionally, there is a brief analysis of the potential market for a new solution in this space.

3.1 Current solutions

Currently, there exist many online platforms for education, some work with schools to aid in traditional learning; others are entirely independent, which students use voluntarily to enhance their skills in a particular topic. Additionally, these independent platforms are not only used by students studying in the field of IT but also by professionals wanting to sharpen their skills or learn a new topic.

Following is a list of some of the best solutions to the problem, and a small summary on them, explaining their pros and cons [13]:

- **Udemy** – This is a paid platform open to the public where anyone can sign-up, the cost can be high, but they often have sales where prices are discounted to very reasonable levels. In terms of quality of content, that is dependent on the instructor, and the course, all of them have star ratings which users can take into account before selecting. This platform is not focused on cybersecurity education, those are just a category of courses, and hence there is no integrated lab environment or a hands-on lab.

- **Open University** – This is also a paid platform which offers educational training in various subjects, even providing accredited qualifications. However, this platform has no specialisation in cybersecurity.

- **Coursera** – This is a somewhat popular platform for distance learning, where various educational institutes mostly design the courses. Hence, it provides a good quality of education, and the pricing is flexible as it is possible to enrol for free

and then pay for certification. However, this suffers from the same issue as Udemy, which is having functional theory but no practical environment to practice.

- **Cybrary** – This is an excellent example of a platform which is focused on cybersecurity education, and indeed this offers courses which apply to various levels of experience and various topics in cybersecurity. However, this platform lacks a lab environment in the free version, in the paid version it offers browser-based virtual labs.

- **Hack The Box** – This is a web platform focused on more experienced users, as even registering for this platform required one to solve a scavenger hunt for the invite code. That is only to allow people with enough experience to join, as this platform focuses only on the practical side of cybersecurity. They have serves and websites which are available for the user to attempt to hack using a variety of skills.

- **Linux Academy** – This is a platform focused on teaching skills in system administration fields including a focus on cybersecurity, and it has very high-quality study literature and also has a playground environment where users can experiment and practice. The challenging part is that it is costly, especially for students who are just starting learning and cannot use this platform to the full potential.

- **Rangeforce** – This is a relatively smaller platform compared to the others; it is used often in Estonian educational institutes in classes which focus on cybersecurity. This platform is a perfect example, offering high-quality, practical content. The most significant limitation for this platform is that it is not open for anyone to join. This platform works by making deals with universities, schools, or companies and offering invite link which are availed by the students or employees, and a casual user cannot go and register to use the platform.

What we learn from the comparisons is that there is need for a platform which can freely and affordably provide high-quality theoretical resources, a lab environment for users to gain practical experience, and an isolated place for users to experiment freely.

3.2 Potential markets for a new cybersecurity platform

The cybersecurity market is growing, it was valued at US$ 104.6 billion in 2017, and it is projected to reach US$258.99 billion by 2025, with a CAGR (Compounded Annual Growth Rate) of 11.9% [34]. More and more companies are now in need of cybersecurity professionals to secure their IT systems from unauthorised use and attacks, especially in the recent times when companies start using cloud services and have work environments with the BYOD (Bring Your Own Device) concept, where users get their own hardware and connect it to the company network to perform work tasks. This concept is especially true for small and medium-scale companies since they work on smaller budgets, and this way of working suits them.

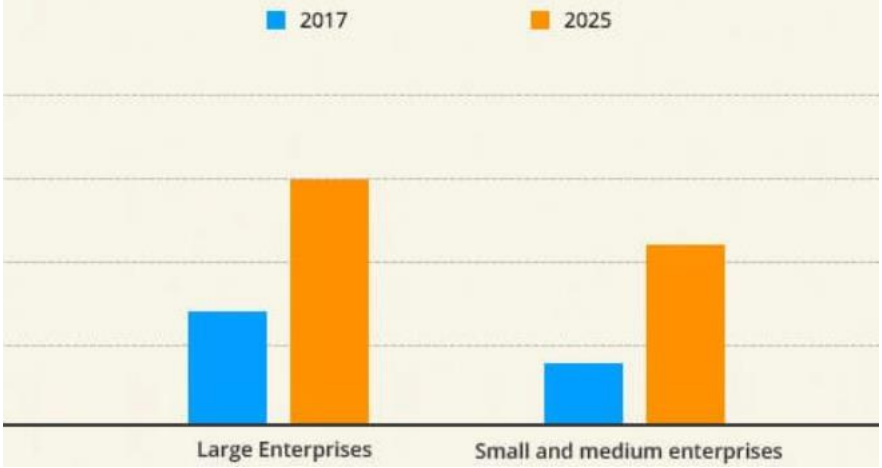

Figure 1 – Global cybersecurity market by user type [34]

This point is further strengthened when looking at Figure 1, where we see that small and medium enterprises will show the highest CAGR in the 2018-2025 period at 14.5% [34]. This data means there will be a high demand for cybersecurity professionals as well as other professionals who have a basic knowledge in the field of cybersecurity. For instance, a developer who has some skills and understandings of cybersecurity concepts is ideally preferred since using that approach the companies can set their entire development process in a way which does not need significant changes later when reviewed by the cybersecurity professionals.

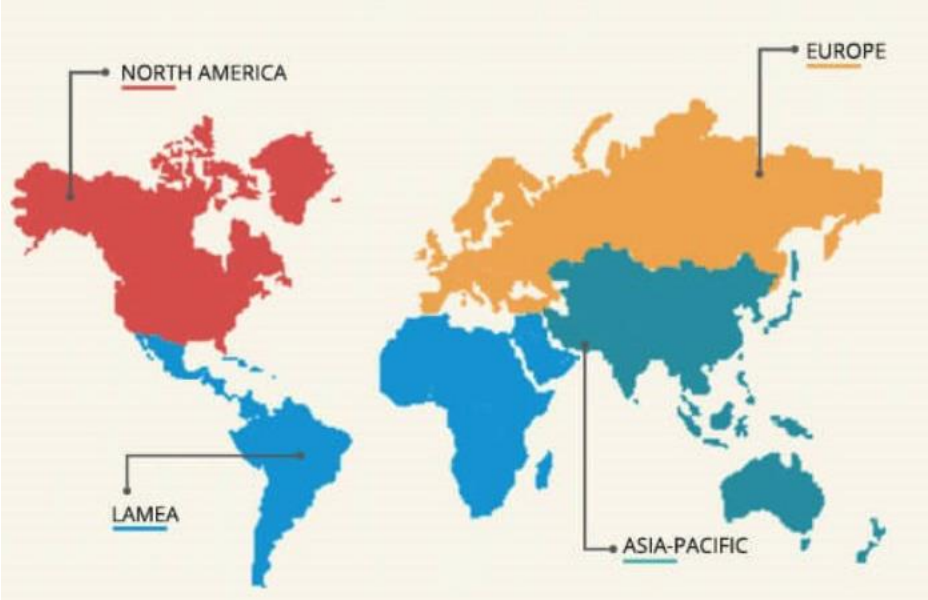

Figure 2 – Global cybersecurity market by region [34]

Globally, it has been established that the demand for cybersecurity will grow, now it should also be researched where exactly most of this growth is expected. The highest CAGR is expected from the Asia-Pacific region at 13.6% [34]. This increased expectation is not by accident; the Asia-Pacific region is having a rapid growth in its economy, as well as countries in the region are being more digital as time proceeds.

A study done by the Centre for Long-Term Cybersecurity at the University of California Berkley concluded with a point stating that unlike in the west, where speed and access were the top priorates in the internet expansion, and security was mainly an afterthought, Asia can shape its cybersecurity at an earlier stage of its internet growth [14]. Keeping that in mind, one country which this solution is going to focus on is – India. It is a country which has a substantial and growing population, a large and growing economy, and it is becoming more and more digital every day. That makes it an attractive target market for such an education platform teaching a skill which is in demand and whose demand will only increase as time goes on. The author completed his school studies in India; and understands the dynamics of education at the school level in the country, and therefore is in a prime spot to incorporate such a platform in the education system which currently is not as focused on cybersecurity education as it should be.

4 Teaching methodologies

Now that the various solutions have been compared, one other thing needs to be analysed – the teaching methodology. This analysis is crucial since the aim of the solution is to provide an educative experience, and it is very different in digital environments compared to the in-person trainings. There are several concepts which this chapter looks into when talking about the teaching methodologies for digital platforms, like:

- E-learning (Electronic learning)
- Self-paced learning
- Lab-based learning

All of them are looked into in the sub-chapters below. Additionally, there is an analysis of a survey created by the author about the effectiveness of these digital teaching methodologies.

4.1 E-learning

E-learning is the concept of taking classes digitally, instead of going to a class and listening to a lecture, in e-learning one usually have live or pre-recorded lectures and online quizzes based on those lectures to evaluate them.

The use of e-learning has been increasing in the recent years, for instance, Tallinn University of Technology, which has taken up platforms like Moodle to provide online resources, quizzes, and notifications, or echo360 to provide lecture recordings online for students to study digitally if necessary.

Based on a study conducted on the effectiveness of an e-learning platform (Moodle in this case) [16] in a university where students stared using the Moodle platform with no prior experience of it or any similar platforms, the statistics showed the following:

Demographics: 80% students aged 21-25

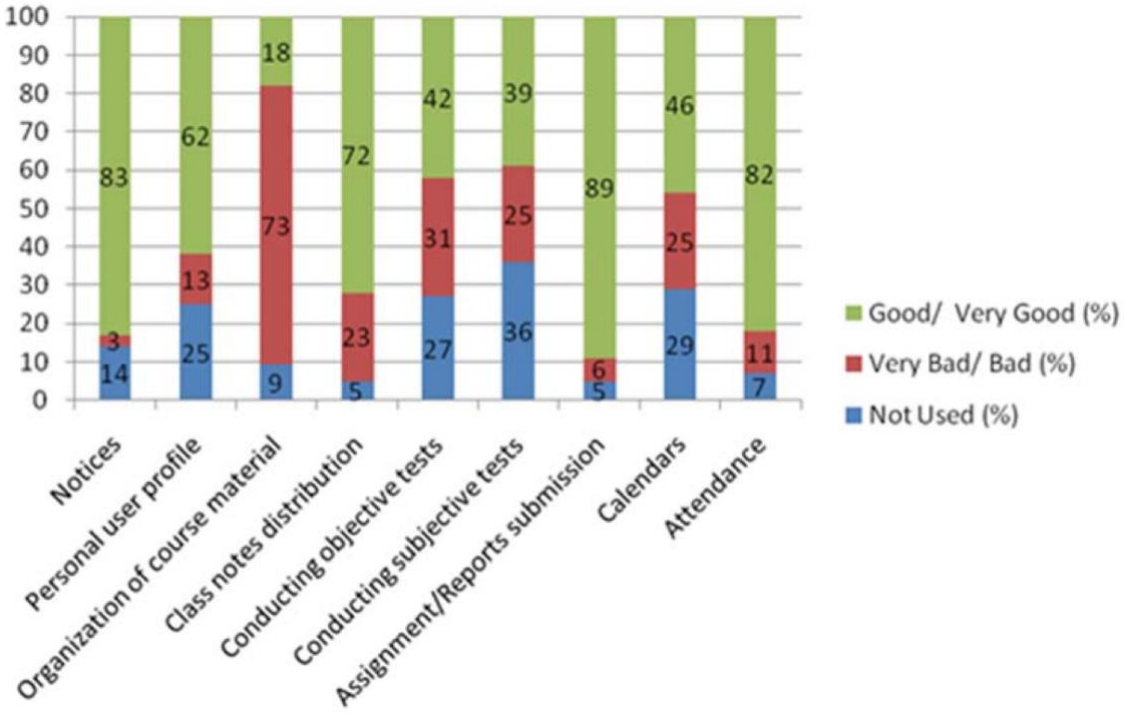

Figure 3 – User experience of Moodle features [16]

In the study, the users were asked to give feedback on different Moodle features, like if they used them and if they had a good or bad experience with it. As seen in the graph, for features like notices, assignment and report submission, and attendance have an overwhelming majority of positive feedback. The most unsatisfactory part in this study was the organisation of course-martial, which since 2015 has improved as the platform has developed. Additionally, the author being a Moodle user himself since 2016 has experienced similar poorly organised courses while in university, however, they were caused in almost all cases by the teacher not having enough knowledge on how to use or organise items on Moodle.

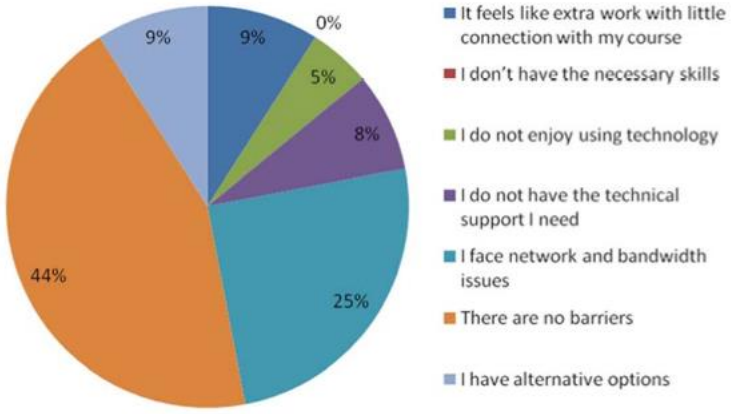

Figure 4 – Moodle barrier to enter pie chart [16]

In the same study, when the students were asked to select anything which they felt as a barrier for using this e-learning platform, the most significant percentage of students said that there were no barriers. The second biggest group had issues which were not directly related to the Moodle platform; they had network and bandwidth issues. The other barriers included issues due to either students having alternatives, finding it too much effort to use, or having issues with technology and in need of support. The issue for networking has now become a non-issue for the most part, since in India, where this survey was, the networking speeds have increased, and the costs have reduced significantly since 2015 ([37][38]). The other challenges of having alternatives, that mostly means attending physical classes, and if a student feels that is ideal and those classes are offered that is not a problem for the e-learning concept. The support needed for such digital-based systems for non-technology friendly users is taken care by giving tutorials for platforms like Moodle during the orientation.

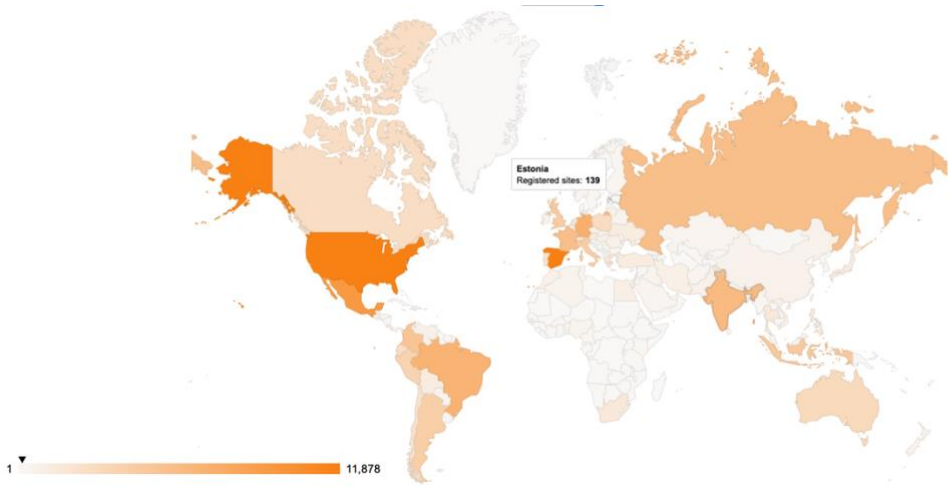

Figure 5 - Moodle global adoption map [36]

This map shows the number of sites running Moodle globally in 2020. This map shows just one e-learning platform; others like Moodle increase the overall number of e-learning platforms even further.

As a conclusion to the study, it was found that the result was that the e-learning platform was effective. It served its purpose and made the overall educational experience better.

Additionally, as it became evident during the 2020 SARS-CoV-2 emergency, digital learning became the only way to continue the school year at all levels of educational institutes. The analysis for how that impacted the education quality shall be only possible after the emergency is over, which at the time of writing is still ongoing. However, from a personal standpoint, for IT education, the form of digital learning is effective.

4.2 Self-paced learning

In many ways a lot of the e-learning platforms, including Moodle in many cases offer self-paced learning, what that means is primarily students learning at their speed. That means there need not be a fixed daily schedule, and the students can learn when they have time and when they feel best suited to start learning new topics. [17]

The benefits self-paced learning provides is that it allows students to learn at their own time and terms. It also has the benefit of allowing people to access course materials at their speed, meaning they can go through topics which they know faster and spend more time on what they need to give extra attention. This way results in lesser wasted time and

promotes efficient learning as the student can manage the learning so that it promotes independent thought and planning. [18]

A research paper on the effectiveness of self-paced learning experimented to test if it is beneficial for learning to self-pace learning. The result of that experiment stated that self-paced subjects revealed significantly higher performance. [19]

4.3 Lab-based learning

Lab-based learning is a method of training where the concepts what a student learns in the theory lessons can be practised in a simulated environment. This method can be applied to many disciplines of learning not only IT, like in business studies many schools organise a mock stock exchange, or a mock supply-chain set-up to teach the students about how those things work. In IT we can take it even further, for simulation one can have one VM (Virtual Machine) to practice something small you learnt, or you can set-up two dozen VMs, to simulate a full IT infrastructure of a company to practice cyber-defence for instance.

In 2010 a research was done into the effectiveness of virtual labs in E-learning, what that research found was that students and teachers knew about virtual labs, and they had a positive feeling towards such an environment. That research concluded by saying that we should make a move to include more virtual lab environments in our teaching system.[20]

4.4 Authors Survey form

The author created a survey asking questions about online education, specifically about e-learning, self-paced learning, and lab-based learning and shared it on social media platforms. [20]

Demographics of surveyors:

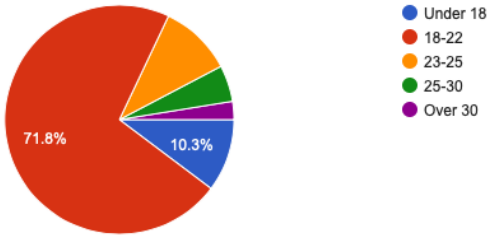

Figure 6 - Age group of surveyors [20]

The split between gender was 46.2% female to 53.8% male, the age group, as shown in Figure 6 was mostly 18-22 years of age, overall the majority of people were under 25 years old. This age group is ideal for surveying about online learning, as most of them are currently in or have completed bachelor's degrees and the second biggest group is of high school students as shown in Figure 7.

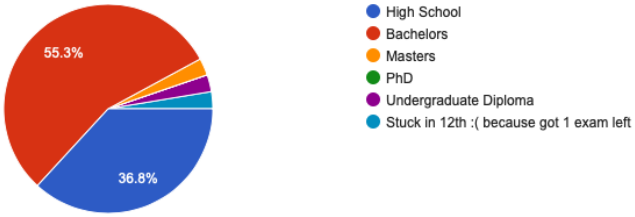

Figure 7 - Current/Completed Study level of surveyors [20]

There were four questions in the survey; one of them was asking the users about what online platforms they have used, if any at all. The others were objective questions, and their results are as follows:

Question: **Have you ever used e-learning platforms? (in school or otherwise)**

The results for this question were 75% of surveyors saying yes, which sets a good understanding of the upcoming questions as most of them are answering from their own experience in that case as they have used these platforms.

Question: Do you know what self-paced learning means? And if yes, do you think is it effective?

Figure 8 - Survey result in asking if self-paced learning is effective [20]

The results in Figure 8 clearly show a majority of responses, knowing what it is and believing that it is effective. Some people answered yes, but with additional comments like in some cases it might work, and in others, it might not, so the results can vary depending on the subject or the student.

Question: Do you believe in a lab-based/ hands-on learning experience? (Where you practice what you learn in theory)

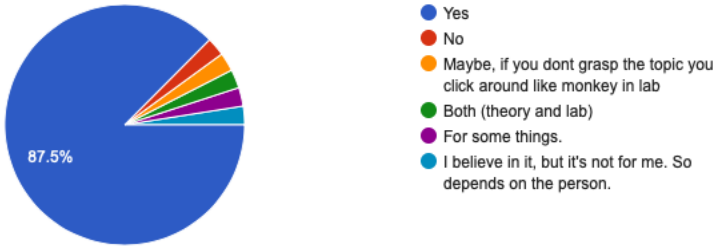

Figure 9 - Survey result in asking if surveyor believed in lab-based learning [20]

The results in Figure 9 show that an overwhelming majority of responses are in full support of virtual lab-based learning. A small percentage of people had an opinion that it might work in some situations, and some say that they do not prefer it.

As a summary to the survey, among the surveyors, it confirms the idea that an e-learning platform which offers self-paced learning and lab-based learning is something in which people show interest. Hence this methodology is correct to go ahead and analyse our solution development and proceeding to develop the solution.

5 Analysis into solution development

In the previous chapters, it has been established that the ideal way to learn independently is in a self-paced way using an e-learning platform, which uses the concept of teaching with the help of virtual lab environments. Hence, in the solution, which is an e-learning platform, is going to incorporate all the analysed teaching methodologies.

The aim of the solution by the end of this thesis is to have a proof-of-concept level of the intended platform running. That includes having at least one training available to show the theoretical side of the platform and an exercise to demonstrate the practical side as well as the automated server deployment process.

In the coming sub-chapters, there is an analysis of the considerations which had to be taken to develop this solution, like choosing the programming language and framework, and the underlying infrastructure elements. All this also takes into account the cost, as the aim is to keep the costs to a minimum as this is a self-funded project by the author.

5.1 Development Language and framework analysis

5.1.1 Language analysis

There are many programming languages which exist for doing web application development, mostly used languages are PHP, Node.js, C#, Python. Each of them has some benefits and some limitations. Below is a brief comparison for the languages.

1. **PHP**:

It has cross-platform compatibility, and is very popular for web apps. It is easy to do templating on it when have prior knowledge of HTML. Additionally, it has a vast developer community, so if one gets stuck, it is easy to find answers. It does have some drawbacks like it is not ideal for desktop apps, and it runs slightly slower compared to some other languages. [21]

2. **Node.js**:

Since Node.js is written in JavaScript, it goes well on both the client and server-side. It can handle the concurrent request with less burden and enables quick loading of a web page. Drawbacks are that error handling and debugging is harder. [22]

3. **C#**:

It is highly scalable, and it is easy to build forms and APIs in it, very useful for cloud-based applications. Drawbacks include a steep learning curve, which means quite a lot of time needed to invest in learning the language. Additionally, it has a small support community compared to PHP, so if one starts to learn it for a substantial project, resources might not be as widely available. [21]

4. **Python**:

It has a straight forward syntax structure, and hence is easier for new learners. It is an object-oriented language similar to the others. It is considered easy to debug, which depends on the developer as it has a unique syntax style, which might be tricky for people migrating from other languages. However, although the syntax is simple for new learners, the learning curve for the language is still steep. [24]

For developing this solution, the decision was made to use PHP for development. This decision is based on the analysis between the languages; additionally one of the most significant factors was considering the previous experience the author had with developing web applications in PHP, as well as the other pros of the language like the broad developer community.

5.1.2 Framework analysis

Now that there is a language selected to develop with, there is a need to select a framework. Frameworks help one develop their application better and faster, as they give certainty that the application is built with a structure and is maintainable and upgradable in the future. Since they let one re-use a lot of the generic code and commonly used modules, it makes the development process faster [28]. Additionally, for security, it is better to use a framework as many services like login, and access controls are included in most of the frameworks, where they are tried and tested by a large community. One can

be confident that these services work correctly; instead of the developer creating. For example, the authentication service is something which certainly is possible to be created by the developer. However, that might increase the possibility of a vulnerability in the application, as the developer could have missed something while programming it, or not tested it properly, or some other unforeseen issue which comes up.

Since the language is PHP, the framework choices have to be made between some of the major PHP frameworks, like Laravel, Symfony, and Yii. The following is a comparison of some pros and cons of each of these frameworks to make the final selection for the solution:

1. **Laravel** [25]:

Pros:

- It stays updated with the newest version of PHP.
- It has a vast ecosystem of additional tools.
- Compatible with other third-party platforms and libraries.

Cons:

- Some applications built in Laravel might be heavier for faster loading on mobile.
- Updates are not compatible; it is possible that if one updates Laravel to a newer version, they could break the code.
- Database migrations are manual.
- Requires decent knowledge of SQL, and usually ties ones app with the database schema design making their application less flexible.

Laravel is ideal if one needs to develop the app quickly and spend less money, as it has an easier learning curve. Advantages with larval is that one can avoid too complicated and too long code. ([25][26])

2. **Symfony** [25]:

Pros:

- A large number of developers are using Symfony actively, and it has one of the most significant communities in the market.
- Symfony is updated regularly to keep it up-to-date with web developers' needs
- Data migration happens automatically, and requires only a simple definition for fields inside the model.
- It does not require significant knowledge of SQL.

Cons:

- It has a steeper learning curve than other PHP frameworks.
- It has a lack of original elements and thus relies on other technologies which could result in longer response times.
- It required more time in testing, and this results in a slower development process.

Symfony is ideal for long-term, complex projects. It has a higher financial toll in case the developers need to be trained on the framework from scratch since it has a steeper learning curve. Although, the higher investment in Symfony pays off in the level of customizability the framework offers. Additionally, Symfony can be faster than the other frameworks if configured correctly. ([25][26])

3. **Yii** [27]:

Pros:

- Rapid development through code scaffolding.
- It has a lot of plugins available.
- It has strong community support.

Cons:

- It can be hard for new users.

- It is not very good for many to many relations.

Yii is ideal as a framework for simple projects which have not much need for customisation. Additionally, it has a steep learning curve. ([26][27])

For this solution, the decision was made to develop using the Symfony framework. Although Symfony has a higher development cost due to a steep learning curve, since the developer of this solution has used Symfony in the past, that cost does not apply in this case. Additionally, since this proof-of-concept solution is going to be upgraded into a more extensive product, it is logical to choose a framework which is stable, fast, secure, and exceptionally flexible which is very important for a platform like this.

5.2 Server deployment analysis

The server deployment analysis is comparing and analysing different software's to be used in the webservers, database servers and hypervisors.

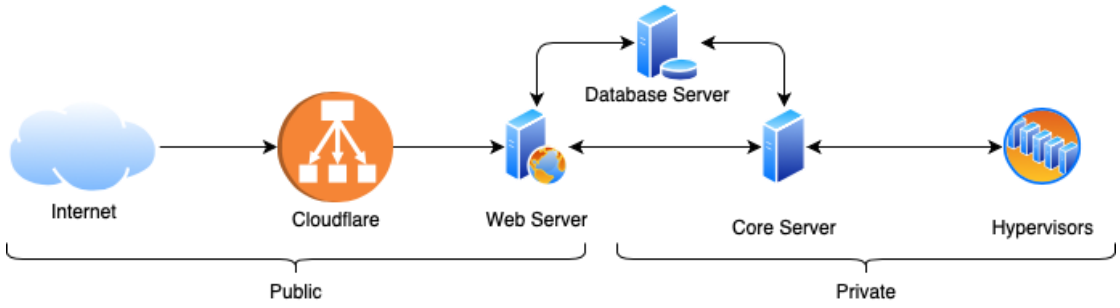

Figure 10 - Solution Infrastructure Diagram (Source: Author created)

Figure 10 visualises the infrastructure architecture of the intended solution quite comprehensively, since the solution is a proof-of-concept, there are not any redundant or load-balancing components; however, that scalability is kept in mind while developing.

The public and private sections of Figure 10 signify the networking between the infrastructural components. Public meaning that that server can be accessed directly over the internet via the URL or IP address. Private meaning that that can only be accessed over a local network; for infrastructure components in different physical locations that translates to having a site-to-site VPN (Virtual Private Network) connection to the other server network. This design enhances security and reduces vulnerabilities, as it significantly limits the publicly accessible infrastructure components.

5.2.1 Web Server

Web servers are used to handle the request the website gets from the users. There are primarily two free web server software's, both of these can be used as a webserver with Symfony framework, but that have their pros and cons, as mentioned below:

1. **Apache** [29]:

   Pros:

   - Very customisable using dynamic modules, and includes 60 official loadable modules.
   - Allows directory level additional configuration using the .htaccess file.
   - More popular than Nginx[30], also has a large community to ask help.

   Cons:

   - The design of this software might demand lots of resources.
   - During heavy load, this could result in severe slowdown due to the architectural design.

2. **Nginx** [29]:

   Pros:

   - Lightweight design allows for faster performance.
   - Can act as a reverse proxy in addition to a web server.
   - Very good for serving static content.

   Cons:

   - Not flexible enough to support dynamic modules and loading, and has third party core modules.
   - It does not process dynamic content natively.

For the solution, the decision was made to use Apache, since the access control offered by the .htaccess file is very useful for having controlled test and development environments on private or publicly accessible networks. Additionally, the author has previous experience of using and configuring Apache with Symfony; Since this software solves the requirement and results in a faster development, it is reasonable to use this.

5.2.2 Database Server

Database servers are used to store all the registered user information, the information on trainings, labs, exercises, and isolated servers which are offered on the platform. In the solution architecture (Figure 10), you see the database server exists in a private network, only accessible to the webserver and core server.

There are many prevalent solutions to use as a database, like MySQL, MariaDB, MSSQL, PostgreSQL, and MongoDB.

In that list of databases, all except MongoDB are RDBMS (Relational database management system), and for our use case we would need to use an RDBMS; hence we shall make a comparison between the remaining four options.

1. **MySQL** [33]:

   Pros:

   - It is very popular.
   - It can very well be configured for high availability use cases.
   - Robust transactional support.
   - If needed, can get enterprise support from a large corporation.

   Cons:

   - Not ideal for large-sized data.
   - Could put a high load on the server if using triggers.
   - Its owned by Oracle, so it not a community-driven open-source project and might be affected by conflicting interests with Oracle’s commercial databases.

2. **MariaDB** [32]:

   Pros:

   - More frequent updates shows an active development community.
   - Highly compatible with MySQL, so easy to migrate from MySQL to this.
   - In use cases like WordPress, MariaDB can provide faster performance.

   Cons:

   - MariaDB is somewhat liable to bloating.
   - Caching is also something which does not work as fast as it can in MariaDB.
   - With newer versions in some cases, it might be needed to recode some part of the application if one migrates from MySQL, as the newer versions are not entirely compatible.

3. **PostgreSQL** [32]:

   Pros:

   - Better performance than others.
   - More flexible compared to others, one can write functions in many server-side languages.
   - It has excellent community support, thanks to its massive user base.

   Cons:

   - The documentation is not the best compared to others.
   - There are better alternatives for performance when using smaller datasets.
   - It does not natively support parallelisation and clustering.

In the solution, the decision was made to use MySQL, since the performance is acceptable for this use-case, since we are not using it to store large-sized data; also, it has excellent support and a large developer community. Additionally, the author has many years of experience with MySQL. Additionally, since the platform is using stock functionality, for the most part, it is easy to migrate to other solutions if it becomes necessary.

5.2.3 Hypervisor software

Hypervisors are servers which are used to deploy virtual machines on servers which have quite a lot of resources for one machine; by virtualising, those resources are allocated to multiple virtual machines (VMs) based on the requirements. Virtualisation is how the multiple virtual labs, exercises, and isolated servers are going to be deployed.

There are many options to use as a hypervisor in this solution like VMware vSphere, Hyper-V, Proxmox PVE and more. Each has its benefits and drawbacks, which we shall see based on the pros and cons list below:

1. **VMware vSphere** [15]:

   Pros:

   - Broad OS support.
   - High-quality support available.
   - Intuitive functionality.
   - It is ideal for large scale infrastructure deployments.

   Cons:

   - Very high cost.

2. **Microsoft Hyper-V** [15]:

   Pros:

   - Simple live migrations.
   - Easy backups.

- Lower priced than VMware vSphere.

Cons:

- It has a limited number of guest OS choices.
- It requires Windows OS during the product lifetime.
- Even though it is cheaper than VMware, it still has a high cost.

3. **Proxmox VE** [23]:

Pros:

- Free to use for personal or business use, and offers optional support plans.
- Debian based, so can use a lot of Linux based commands and features on it.
- Accurately documented REST API.

Cons:

- It does not have the same level of enterprise support as other expensive alternatives.
- It does not have out-of-the-box support for automation tools like Terraform.

In the solution, the decision was made to use Proxmox VE since it is highly customisable and reliable, it can be clustered if needed, and offers excellent support through the community for any familiar problem. Additionally, the author has administered Proxmox VE and developed on its REST API in the past, which is also a factor; however, the most significant factor is the cost, since it is free to use even for commercial use, it is ideal for this solution.

5.2.4 Proxy solution

As a proxy solution, the application is using Cloudflare, as it is a well-reputed proxy solution provider and works as intended. Additionally, one of the most significant factors for this proof-of-concept is that it has a free plan which can be used and still provides necessary features.

6 Development of solution

For the development, the solution was split into smaller parts based on the complexity. These parts are:

- Web application development – this comprises of the platform which the user sees and interacts with; the landing page and the login form are among a few pages part of this section.

- Core application development – this is an application server for the platforms' virtualisation service, this part deals with the process of creating and destroying VMs in the hypervisors and entering the IP addresses and details of issues VMs to the database to keep track of what machine is issued to which user.

- Virtual Machine automation – this deals with the communication between the core and hypervisor, so how the actual creation and deletion of exercise or lab instances is taking place. Additionally, this also deals with the process of deploying isolated VMs to users where they can practice their skills freely, without the threat of breaking anything.

6.1 Web application development

The development of this part is done in PHP using the Symfony framework, as decided during the analysis. In the solution this is the infrastructure element which is located on the border of the public and private networks, the users connect publicly to it, and all other communication it has with other infrastructure elements is happening in a private network.

There are different use cases in which the user uses the web application, which defines how it interacts and communicates with other infrastructure components.

6.1.1 Use case: No virtualisation

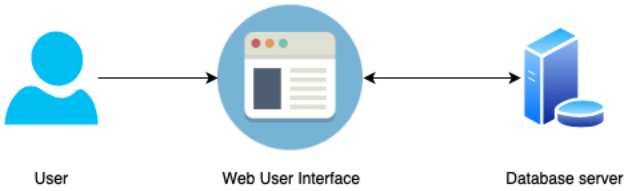

Figure 11 - Web application no virtualisation use case (Source: Author created)

This use case is where the user is interacting with the platform in such ways which do not require any request to be sent to the core server and hypervisors. These interactions include actions like login, user registration, editing the user profile, or accessing the training resources.

6.1.2 Use case: Virtualisation

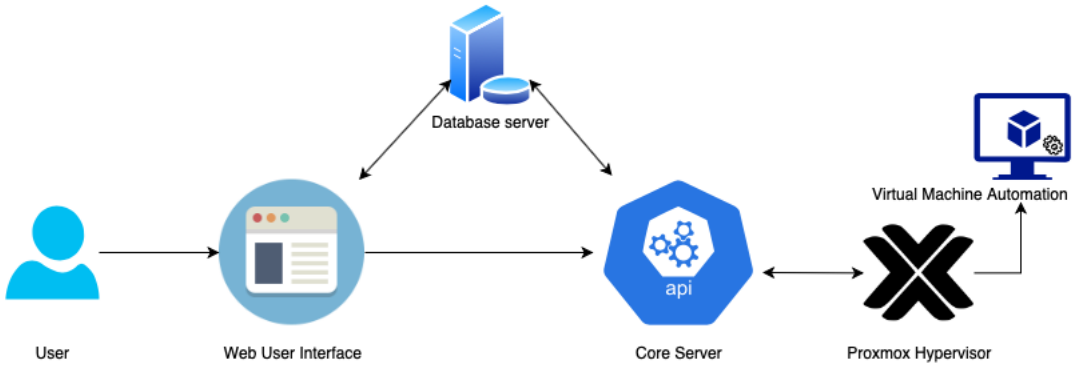

Figure 12 - Web application virtualisation use case (Source: Author created)

This use case is where the user performs actions which require some actions to be performed related to a virtual machine. These can be user launching or stopping a virtual lab or exercise, or requesting, editing or stopping an isolated server.

In all these situations, the web application performs an action in the specific database table, like adding the user and exercise details so that there is a record of who is requesting for what exercise, and then it makes a request to the core server, which is an API (Application Programming Interface) which deals with the communication between the application and hypervisor.

After receiving a response from the hypervisor, the core server updates the database with the IP address and other networking of this exercise server. Additionally, it also adds some server details to the database record like in case of multiple hypervisor nodes it

notes which exact node is this machine located on, this information is useful for the decommissioning of the instances.

6.2 Core application development

The core application is an API which acts as an intermediary between the hypervisor and the web application. It is developed in PHP using the Symfony framework, similar to the web application, the main difference being the core has no web interface, it is meant only to receive and process request like launching or stopping an exercise instance.

The tasks which the core sever performs could have been incorporated in the web application, hence reducing some of the complexity at least at the proof-of-concept level; however, this was developed with scaling and future development in mind. Additionally, keeping in mind the future development, it would be wise to have them separated, as after all their functionality differs significantly. Additionally, when both the web application and the core are fully developed, the codebases of each of them shall be enormous, hence separating them is wise for maintenance purposes as well.

6.3 Virtual Machine automation

This part of the platform consists of the actual deployment process for the virtual labs, exercises, and isolated VMs. The current solution has an exercise, which is used to demonstrate the proof-of-concept, so in this subchapter, the process for deployment for that exercise is described.

To deploy the exercise instance, the core application runs decision logic, and selects the name and IP address, in the current version since it is a proof-of-concept that happens on a simple incremental logic, meaning the current address plus one is used. Once all the details are ready, the core server creates a connection to the Proxmox node. Once that is established, it initiates an API call to the Proxmox node to create an instance, passing all the necessary details. In case of having multiple hypervisor nodes, the core server runs a function to find out the node with the lowest latency to the user and has enough resources to handle the exercise instance.

The calls done to the Proxmox node are done using a PHP library[6] for Proxmox VE, the benefit of this library is that in the code Proxmox nodes are treated as objects which

simplifies development, as well as the passing of arguments for the creation or deletion, is straightforward. The library is open-source with an MIT Licence, which means the platform can use it without restriction. Additionally, the library itself uses Proxmox API, and is stable, working on any PHP version higher than 5.5.

7 Summary

Lack of online resources for cybersecurity training and practice is an actual concern currently. In this thesis it was established that there is a need for a new online platform for cybersecurity learning and training, as platforms currently available limit the students to either a theoretical only learning experience or an expensive platform which has some of the practical options. Platforms with good virtual lab-based trainings are mostly unavailable openly, as they require contracts with the schools or companies.

The goal of the work was to analyse and develop a proof-of-concept for a web-based platform for cybersecurity learning and practicing, which can be made available to anyone who wishes to start learning. The author analysed the current solutions in this space, the teaching methodologies used for digital learning, all of which were used to design the solution.

When analysing the developing process, there were several decisions made to decide on subjects like choosing the programming language, and framework, and choosing the software used as a web server, database server, and hypervisor.

The decisions of those choices were:

- Programming language: PHP
- Programming framework: Symfony
- Webserver software: Apache
- Database software: MySQL
- Hypervisor solution: Proxmox VE
- Proxy solution: Cloudflare

The requirement for the solution was to demonstrate the basic functioning of the final version of the platform, which included having at least one training material which was the theoretical part, and at least one exercise to demonstrate the practical part as well as the automation on the hypervisor side.

Based on the decisions and requirements, the solution was developed, and is now public at the link https://hackstud.io. The platform is divided into multiple sections to make it easy to use, develop, and maintain. The sections are separated based on the services that they offer, such as the training materials section which contains theoretical resources. The virtual labs' section, which consists of labs based on the pieces of training. The exercise section, which consists of challenges designed to let users practise their skills in the field of cybersecurity. The sandbox section, which deals with the deployment of isolated servers. These isolated servers are there for users to experiment with their skills without the risk of damaging any other systems, as these are completely isolated on the network.

The solution met the requirements set for it, having one theoretical training and one exercise option available. The topic of theoretical resource used as an example is 'How to set a static IP in Linux', this was chosen as the topic since it is a fundamental task which is often used by IT students and serves as a good example to demonstrate the platform's theoretical side. For the practical demonstration of the platform, an exercise was created which deploys an instance of 'OWASP JuiceShop' which the users can practice on; this was chosen as the OWASP (Open Web Application Security Project) is a large online community in the field of web application security. Additionally, the 'JuiceShop' exercise is a comprehensive challenge demonstrating a variety of cybersecurity vulnerabilities, which makes it an excellent resource for the users to practice.

This thesis established why we need a web-based platform available to the public openly for learning and practicing cybersecurity, and successfully demonstrated a solution at the proof-of-concept level.

What makes the solution ideal is the fact it offers both a theoretical side and a practical side, all in one platform. So the user can start learning a concept, practice it in a lab, and launch an exercise to get better at the skills or experiment more. Additionally, beyond just that, for users wanting to get a server which is completely isolated from others to

practice without the risk of breaking anything can do so in the same platform. During the research, there was no other platform which offered such a comprehensive solution to the problem. Especially in the primary focus market of India, there are very few platforms even offering virtual labs or exercises, nowhere near the comprehensive solution this platform offers.

Future work plans on this platform include improving the platform in terms of the user interface as well as adding more functionality, and adding more resources to the platform such as more exercises, more trainings with related labs, and making the isolated servers available for the users to order. Another main focus for future development shall be to include trainings, virtual labs, and exercises on networking specific topics.

References

[1] Dictionary Cyber Security definition Weblink: https://www.dictionary.com/browse/cybersecurity?s=t [Last Accessed: 11/04/2020]

[2] The 15 biggest data breaches of the 21st century. Weblink: https://www.csoonline.com/article/2130877/the-biggest-data-breaches-of-the-21st-century.html [Last Accessed: 25/04/2020]

[3] Zoom grows to 300 million meeting participants despite security backlash. Weblink: https://www.theverge.com/2020/4/23/21232401/zoom-300-million-users-growth-coronavirus-pandemic-security-privacy-concerns-response [Last Accessed: 26/04/2020]

[4] Zoom privacy and security issues: Here's everything that's wrong (so far). Weblink: https://www.tomsguide.com/news/zoom-security-privacy-woes [Last Accessed: 28/04/2020]

[5] Zoom security issues: Zoom could be vulnerable to foreign surveillance, intel report says. Weblink: https://www.cnet.com/news/zoom-every-security-issue-uncovered-in-the-video-chat-app/ [Last Accessed: 30/04/2020]

[6] Proxmox VE API Client. Weblink: https://github.com/ZzAntares/ProxmoxVE [Last Accessed: 30/04/2020]

[7] Zoom has added more videoconferencing users this year than in all of 2019 thanks to coronavirus, Bernstein says. Weblink: https://www.cnbc.com/2020/02/26/zoom-has-added-more-users-so-far-this-year-than-in-2019-bernstein.html [Last Accessed: 27/04/2020]

[8] Bangladesh Bank Heist. Weblink: https://youtu.be/Usu9z0feHug [Last Accessed: 27/04/2020]

[9] That Insane, $81M Bangladesh Bank Heist? Here's What We Know. Weblink: https://www.wired.com/2016/05/insane-81m-bangladesh-bank-heist-heres-know/ [Last Accessed: 28/04/2020]

[10] The Billion Dollar Bank Job. Weblink: https://www.nytimes.com/interactive/2018/05/03/magazine/money-issue-bangladesh-billion-dollar-bank-heist.html [Last Accessed: 28/04/2020]

[11] Bangladesh hack illustrates rising sophistication of attacks. Weblink: https://assets.kpmg/content/dam/kpmg/xx/pdf/2016/08/swift-it.pdf [Last Accessed: 28/04/2020]

[12] Bangladesh Bank Heist: Lessons Learned. Weblink: https://www.bankinfosecurity.com/bangladesh-bank-heist-lessons-learned-a-9064 [Last Accessed: 28/04/2020]

[13] Best online cybersecurity courses of 2020: free and paid certification programs, degrees and masters. Weblink: https://www.techradar.com/best/best-online-cyber-security-courses [Last Accessed: 24/04/2020]

[14] 2020 Roundup Of Cybersecurity Forecasts And Market Estimates. Weblink: https://www.forbes.com/sites/louiscolumbus/2020/04/05/2020-roundup-of-cybersecurity-forecasts-and-market-estimates/#7acd2369381d [Last Accessed: 30/04/2020]

[15] Hyper-V vs. VMware: Which Is Best? Weblink: https://www.atlantech.net/blog/hyper-v-vs.-vmware-which-is-best [Last Accessed: 30/04/2020]

[16] EFFECTIVENESS OF MOODLE-ENABLED BLENDED LEARNING IN PRIVATE INDIAN BUSINESS SCHOOL TEACHING NICHE PROGRAMS. Weblink: http://tojned.net/journals/tojned/volumes/tojned-volume05-i02.pdf#page=20 [Last Accessed: 27/04/2020]

[17] Self-Paced Distance Learning: How Does IT Work? Weblink: https://study.com/articles/Self-Paced_Distance_Learning_How_Does_it_Work.html [Last Accessed: 27/04/2020]

[18] What's so good about self-paced learning?. Weblink: https://www.nickelled.com/blog/whats-so-good-about-self-paced-learning/

[19] On the effectiveness of self-paced learning. Weblink: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3079256/

[20] Survey form for online education Weblink: apd.ee/thesis-survey [Last Accessed: 29/04/2020]

[21] The Pros and Cons of PHP and .NET. Weblink: https://www.informaticsinc.com/blog/september-2018/pros-and-cons-php-and-net [Last Accessed: 27/04/2020]

[22] Pros and Cons Compared: Node.js vs. Python. Weblink: https://www.ongraph.com/pros-cons-compared-node-js-vs-python/ [Last Accessed: 27/04/2020]

[23] Proxmox vs. VMware vs. Cloud. Weblink: https://sweetcode.io/proxmox-vs-vmware-vs-cloud/ [Last Accessed: 30/04/2020]

[24] Developing Solutions in NodeJS vs Python: Pros and Cons. **Weblink**: https://medium.com/@Intersog/developing-solutions-in-nodejs-vs-python-pros-and-cons-4ab4cea68ff0 [Last Accessed: 27/04/2020

[25] Laravel vs Symfony in 2020 – which framework choose for your project?. **Weblink**: https://asperbrothers.com/blog/laravel-vs-symfony/ [Last Accessed: 27/04/2020]

[26] How to choose a PHP framework. **Weblink**: https://opensource.com/business/16/6/which-php-framework-right-you [Last Accessed: 27/04/2020]

[27] Yii vs Symfony detailed comparison. Weblink: https://www.slant.co/versus/3757/3758/~yii_vs_symfony [Last Accessed: 27/04/2020]

[28] Why should I use a framework?. Weblink: https://symfony.com/why-use-a-framework [Last Accessed: 27/04/2020]

[29] Apache Vs NGINX – Which Is The Best Web Server for You?. Weblink: https://serverguy.com/comparison/apache-vs-nginx/ [Last Accessed: 28/04/2020]

[30] Comparison of the usage statistics of Apache vs. Nginx vs. Microsoft-IIS for websites. Weblink: https://w3techs.com/technologies/comparison/ws-apache,ws-microsoftiis,ws-nginx [Last Accessed: 28/04/2020]

[31] MariaDB Vs MySQL In 2019: Compatibility, Performance, And Syntax. Weblink: https://blog.panoply.io/a-comparative-vmariadb-vs-mysql [Last Accessed: 28/04/2020]

[32] Comparing 3 open source databases: PostgreSQL, MariaDB, and SQLite. Weblink: https://opensource.com/article/19/1/open-source-databases [Last Accessed: 28/04/2020]

[33] MariaDB vs MySQL: Key Performance Differences. Weblink: https://www.guru99.com/mariadb-vs-mysql.html [Last Accessed: 28/04/2020]

[34] Cyber Security Market Outlook – 2025. Weblink: https://www.alliedmarketresearch.com/cyber-security-market#toc [Last Accessed: 30/04/2020]

[35] Zoom Security Issues. Weblink: https://youtu.be/QQ7mSpVyIJo [Last Accessed: 26/04/2020]

[36] Moodle Statistics. Weblink: https://stats.moodle.org [Last Accessed: 26/04/2020]

[37] Cisco: VNI Complete Forecast Highlights - India Weblink: https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/India_2020_Forecast_Highlights.pdf [Last Accessed: 18/05/2020]

[38] The Economics Behind India’s Super-Cheap ($0.26 Per GB) Mobile Data. Weblink: https://medium.com/swlh/the-economics-behind-indias-super-cheap-0-26-per-gb-mobile-data-40f28bdd7774 [Last Accessed: 18/05/2020]