TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Mykyta Zaitsev 184676IVCM

Developing Learning Pathways for Massive Open Online Courses by the Example of the RangeForce CyberSkills Training Platform

Master's Thesis

Supervisor
Olaf Manuel Maennel
Professor

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:	Mykyta Zaitsev	.... (signature)
Date:	May 10, 2020	

Abstract

Massive Open Online Courses (MOOC) have grown steadily over the past two decades. Over this time, the largest players on the market have experienced multiple problems such as learners' disorientation and low engagement. To solve this problem innovative MOOCs adopted the methodology of learning pathways. Designed learning pathways solved the disorientation problem by directing learners step by step, while the executed learning pathways helped to analyze personal learning styles, performance, and engagement. However, not every MOOC has a preset learning pathway, which can potentially harm learning experience. But what exactly happens if a MOOC does not have learning pathways in place? I carry out a complex research of an existing MOOC without learning sequence to find out that non-directed learner behavior is irrational and unproductive. Based on that analysis, I describe how transition from unstructured learning sequence to a learning pathway-based educational environment can benefit MOOCs. I also present a solution for the MOOC in question by developing two learning pathways after manually evaluating all MOOC's modules and their application to the NIST-NICE framework.

The thesis is in English language and contains 41 pages of text, 5 chapters, 32 figures, 3 tables.

List of abbreviations and terms

MOOC	Massive Open Online Courses
OER	Open Educational Resources
CCK08	Connectivism and Connective Knowledge
MIT	Massachusetts Institute of Technology
RSS	Really Simple Syndication
ANOVA	Analysis of Variance
RF	RangeForce
WASE	Web Application Security Essentials
SOC	Security Operations Center
HTTPS	HyperText Transfer Protocol Secure
NIST	National Institute of Standards and Technology
NICE	National Initiative for Cybersecurity Education
CIR	Incident Response Specialty Area
EXP	Exploitation Analysis Specialty Area
VTA	Virtual Teaching Assistant
IDS	Intrusion Detection System
IPS	Intrusion Prevention System
CSCUK	Cyber Security Challenge UK
MS	Microsoft
PCAP	Packet Capture
CTF	Capture The Flag

Table of Contents

List of Figures vi

List of Tables viii

1 Introduction 1

1.1 Main Problems . . . 1

1.2 Main Objectives . . . 1

1.3 Acknowledgements . . . 2

2 Online Learning in Educational Technology 3

2.1 Emergence of MOOCs . . . 3

2.2 Concept of the Learning Pathway . . . 5

2.3 Learning Pathways in MOOCs . . . 7

3 Structure of Learning Pathways 9

3.1 Designed v. Executed Learning Pathways . . . 9

3.2 Designed Learning Pathways . . . 10

3.2.1 Learning Outcomes . . . 10

3.2.2 Strategic Prompts . . . 10

3.2.3 Connections Between Stops . . . 11

3.2.4 Scalability of Modules . . . 12

3.2.5 Visual Representation of Learning . . . 12

3.3 Executed Learning Pathways . . . 12

3.3.1 Tutoring System . . . 13

3.3.2 Data Collection and Metrics . . . 14

3.3.3 Patterns Discovery . . . 15

4 Analysis of Learning Behavior 17

4.1 Research Subject . . . 17

4.2 Analysis of Executed Learning Pathways . . . 19

4.2.1 Ethical Considerations . . . 19

4.2.2 Data Collection and Metrics . . . 20

4.2.3 Visualization and Patterns Discovery . . . 22

4.2.4 Outtakes . . . 27

5 Development of Designed Learning Pathways 28

5.1 Development of Designed Learning Pathway 28
5.1.1 Connections Between Stops 29
5.1.2 Learning Outcomes 32
5.1.3 Scalability of Modules 34

6 Summary **35**

Bibliography **36**

Appendices **38**

Appendix 1 - RangeForce modules as of February 5, 2020 **38**

List of Figures

1 *MOOCs and open-education timeline, retrieved from http://blogs.cetis.org.uk/* 4
2 *Cousera's career learning paths, retrieved from https://blog.coursera.org/* 7
3 *Cousera's Data Scientist learning path, retrieved from https://careers.coursera.org/data-scientist/* 8
4 *Difference between designed (left) and executed (right) learning pathways, retrieved from https://chauff.github.io/* 9
5 *Coursera's learning outcomes are represented as acquisition of career skills, retrieved from https://www.coursera.org/* 10
6 *Example of backtracking can be visible in the middle of Weeks 2 and 3, retrieved from http://www.educationaldatamining.org/EDM2016/* 11
7 *Examples of Stops' Placements* 11
8 *Khan Academy's Dashboard, retrieved from https://www.khanacademy.org/* 12
9 *Executed pathways of non-passing and passing learners, adapted from https://chauff.github.io/* 13
10 *Metrics of H. Chuang and C. Shen's experiment, adapted from "A Study on the Applications of Learning Paths Conceptsto the Teaching in Elementary School", by H. Chuang and C. Shen, 2008.* 14
11 *10 motifs from research on Wayang Outpost, adapted from "Identifying High-Level Student Behavior Using Sequence-based Motif Discovery", by by Shanabrook, Cooper, Woolf and Arroyo, 2010.* 14
12 *Distribution of Learning pathways, adapted from "A Study on the Applications of Learning Paths Concepts to the Teaching in Elementary School", by H. Chuang and C. Shen, 2008.* 16
13 *Descriptive statistics of learning outcome in different learning path, adapted from "A Study on the Applications of Learning Paths Concepts to the Teaching in Elementary School", by H. Chuang and C. Shen, 2008.* . . . 16
14 *View of RangeForce Userhub, retrieved from https://hub.rangeforce.com/learn* 18
15 *Virtual Teaching Assistant 1.4 is the tutoring system of RangeForce, retrieved from https://hub.rangeforce.com/learn* 19
16 *Pseudonymized usernames in users table, example in R language.* 19
17 *Log entries by RF-related users and learners as of Februrary, 5, 2020* . . 20
18 *All extracted module attempts of UserID 2, example in R language.* 21
19 *Representation of users' next steps after completing modules with the progress higher than zero, example in R language.* 21

20 *Extraction of vertices with the highest progress, example in R language.* . 22
21 *Script used for visualization of the dataframe, example in R language.* . . 22
22 *Visualization of module relations which take origin from DevOps topic.* . . 23
23 *Examples of promo codes which involve DevOps - Introductory Lab.* 24
24 *Visualization of module relations which take origin from the WASE topic.* . 24
25 *Visualization of module relations which take origin from the SOC topic.* . 25
26 *Visualization of module relations which take origin from Security Tools topic.* 26
27 *Top 10 users by the number of module attempts, left column - pseudonymized id, right column - number of module attempts.* 27
28 *Top 10 modules by the number of attempts, left column - module id, right column - number of module attempts.* 27
29 *Transition of the initial model of DevOps topic to the linear model* 30
30 *Ordered lists of topics: blue - DevOps, red - Security Tools, orange - SOC, green - WASE* 31
31 *Mapping of RangeForce modules to NIST-NICE's Specialty Areas* 32
32 *Incident Response (CIR) and Exploitation Analysis (EXP) learning pathways* 34

List of Tables

1 *A table of implemetation of the learning pathways in biggest xMOOCs* . . 8

2 *Table of filtered out modules by category* 28

3 *Descriptions of NIST-NICE Specialty Areas: Incident Response (CIR) and Exploitation Analysis (EXP)* 33

1. Introduction

This research overviews current situation in Massive Open Online Courses. MOOCs tend to have plenty of training content. Once a MOOC acquires a vast library of materials, it becomes hard to navigate through. As a result, without properly devised learning assistance, disorientations and inefficiencies occur, which later result in significant deviations from originally established teaching objectives, low engagement, and consequently lead to high dropout rates. Some MOOCs have implemented learning sequences to mitigate previously mentioned disadvantages. Currently, there are still MOOCs that have not implemented them yet, and these MOOCs might face negative consequences. The first goal of this research is to analyze a real-life case of a MOOC without a curriculum in place in order to evaluate diverse user behaviors. The prove of high diversity and irrationality of learners' behaviour is the root cause for implementation of a learning sequence for the MOOC. Another goal of this research is to develop a learning sequence based on methodology of learning pathways.

1.1 Main Problems

The existing MOOCs have integrated network-based teaching resources, resource management, and digital supplementary materials. However, some existing online education systems have a common problem: they do not provide possibilities of preset learning pathways, which would enable users to have a clear understanding of how their learning process should be organized. Different users of the same MOOC generally show different learning behaviors based on their personal abilities, interests and demands. As a result, without properly devised learning assistance and help, disorientations and inefficiencies occur, which later result in significant deviations from originally established teaching objectives, low engagement, and high dropout rates. [1]

1.2 Main Objectives

This study focuses on learning sequences in MOOCs. My aim is to research the downsides of the absence of learning directions and continue with the concept of learning pathways and their benefits. This study holds an experiment on a cyber security-aimed MOOC with an absent learning path with a purpose of designing one.

The primary outcome of this thesis will be:

Objective 1. Analysis of the relevant literature and use-cases on learning pathways and the problems they solve.

Objective 2. Analysis of *learners' behavior* by collecting the log entries of their performance from a cyber security-aimed MOOC.

Objective 3. Developed *designed* learning pathways for a cyber security-aimed MOOC.

This research benefits in multiple ways. It proves that the absence of learning sequences leads to student disorientation and low engagement. It also provides a documented way of creation of a learning sequence following the methodology of the designed learning pathways.

1.3 Acknowledgements

I would like to express my special thanks of gratitude to Margus Ernits, the CTO at RangeForce, for his contribution and opportunity to conduct an experiment using the RangeForce data. Also, I would like to thank Tiia Tänav, a Software Developer at RangeForce, for helping me extract, pseudonymize, and filter that data.

2. Online Learning in Educational Technology

In this chapter the development of the domain of the online learning over the past two decades is being discussed. This chapter reviews the emergence of the open educational resources (OER) movement and its transition into MOOCs. Also, it covers the methodology of learning pathways, which has been developed as a solution for companies wishing to educate their employees, but eventually became an essential part of successful MOOCs. [2]

2.1 Emergence of MOOCs

The first MOOCs emerged from the OER movement, which was sparked by MIT OpenCourseWare[1] project in 2001.

Within the OER movement Wikiversity[2] was founded in 2006. The primary goals of Wikiversity were to:

Goal 1. Create and host a range of free content, multilingual learning materials, for all age groups in all languages.

Goal 2. Host learning projects and communities that support these materials.

The first open course on the platform was organised in 2007. A ten–week course with more than 70 students was used to test the idea of making Wikiversity an open and free platform for education. On Wikiversity learning was facilitated through collaboration on projects that are detailed, outlined, summarized or results reported by editing wiki webpages of a particular topic. Wikiversity participants were encouraged to express their learning goals, while the Wikiversity community collaborated to develop learning activities and projects to accommodate those goals.

The term MOOC itself was coined in 2008 by Dave Cormier of the University of Prince Edward Island in response to a course called *Connectivism and Connective Knowledge* (also known as CCK08). CCK08, which was led by George Siemens of Athabasca University and Stephen Downes of the National Research Council, consisted of 25 tuition-paying

[1] MIT OpenCourseWare `https://ocw.mit.edu/index.htm`
[2] Wikiversity `https://www.wikiversity.org/`

students in Extended Education at the University of Manitoba, as well as over 2200 online students from the general public who paid nothing. All course content was available through RSS feeds, and online students could participate through collaborative tools, including blog posts, threaded discussions in Moodle, and even Second Life meetings.

As MOOCs developed with time, multiple conceptions of the platform seem to have emerged. Mostly two different types can be differentiated: those that emphasize a connectivist philosophy, and those that resemble more traditional courses. To distinguish the two, several early adopters of the platform proposed the terms **"cMOOC"** and **"xMOOC"**:

- cMOOC instructional design approaches attempt to connect learners to each other to answer questions or work together on joint projects. cMOOCs are based on principles of connectivist pedagogy, indicating that material should be aggregated, rather than pre-selected, remixable, re-purposable, and feeding forward.
- xMOOCs have a much more traditional course structure. They are characterized by a specified aim of completing the course obtaining certain knowledge certification of the subject matter. Typically, they are presented with a clearly specified syllabus of recorded lectures and self-evaluation tools. Some providers required paid subscriptions for acquiring graded materials and certificates.

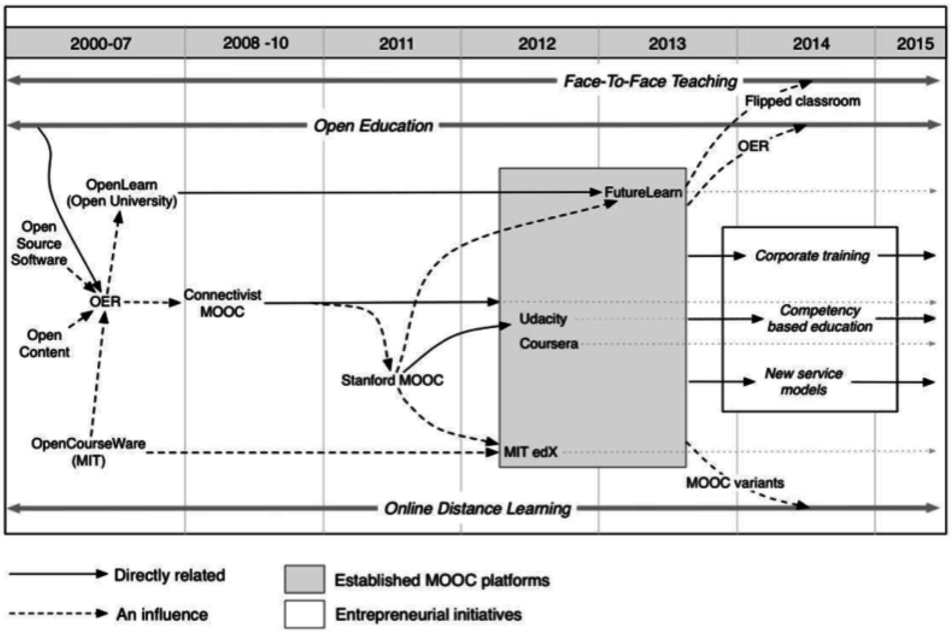

Figure 1. *MOOCs and open-education timeline, retrieved from http://blogs.cetis.org.uk/*

In the fall of 2011, Stanford University launched three courses. The first of those courses was Introduction Into AI, launched by Sebastian Thrun and Peter Norvig. Enrollment quickly reached 160,000 students. The announcement was followed within weeks by the launch of two more MOOCs, by Andrew Ng and Jennifer Widom. Following the publicity and high enrollment numbers of these courses, Thrun started a company he named Udacity and Daphne Koller and Andrew Ng launched Coursera.

Later, according to The New York Times, 2012 became "the year of the MOOC" as several well-financed providers, associated with top universities, emerged, including Coursera[3], Udacity[4], and edX[5]. [3] Many universities scrambled to join in the "next big thing", as did more established online education service providers. Dozens of universities in Canada, Mexico, Europe and Asia have announced partnerships with large American MOOC providers.

By early 2013, questions emerged about whether academia was "MOOC'd out." This trend was later confirmed in continuing analysis. [4]

In January 2013, Udacity launched its first MOOCs-for-credit, in collaboration with San Jose State University. In May 2013 the company announced the first entirely MOOC-based master's degree, a collaboration between Udacity, ATT and the Georgia Institute of Technology, costing $7,000, a fraction of its normal tuition. [5]

Concerned about the commercialization of online education, MIT created the not-for-profit MITx. Later Harvard, University of California, Berkeley, the University of Texas System, Wellesley College and Georgetown University joined the initiative, which was also renamed to edX.

As of January 2016, edX offered 820 courses, Coursera offered 1580 courses and Udacity offered more than 120 courses. According to FutureLearn, the British Council's *Understanding IELTS: Techniques for English Language Tests* had an enrollment of over 440,000 students. [6]

2.2 Concept of the Learning Pathway

Meanwhile, Jim Williams and Steve Rosenbaum, researched the orientation programs in large companies. Some of these programs were designed as boot camps with the intent to immerse new employees in all the training needed to perform their job. The problem in many cases was that the new employees had forgotten much of what was taught in the first few days of training by the time they reached "graduation" and they had to continue to learn and re-learn after training was supposedly completed.

[3] Coursera `https://www.coursera.org/`
[4] Udacity `https://www.udacity.com/`
[5] edX `https://www.edx.org/`

Other companies took a much more practical approach by developing customized orientation and training development plans that mapped out the training and activities that new employees must complete to be considered completely competent at their job. This was closer to an even more effective approach to getting employees up to speed.

As a solution Williams and Rosenbaum proposed the methodology of a learning pathway in their book "Learning Paths: Increase Profits by Reducing the Time It Takes Employees to Get Up-to-Speed" published in 2004. [2]

The solution was a practical approach to defining proficiency and then producing an effective sequence of training, practice, coaching, and experience to accelerate the time it takes for a new employee to reach that defined proficiency. Within that solution it was essential that:

- A Learning Path is defined as a sequence of learning events. It is most effective when it includes opportunities to observe or shadow an employee and then to practice the task.
- A Learning Path is designed in a way that new employees start with executing simpler tasks and continue with the more complex ones as they move forward.
- Proficiency is reached when employees are independently productive, which means that they can complete tasks without assistance and errors.

As an integral part of the Learning Path, technology can be useful in creating practice environments. But, the expense and time invested in developing such environments could pose as an obstacle and the investment, thus, would need to be justified. Technology simulations exist in many forms and are not limited to just software application simulations; there are simulation vendors that provide tools and services to develop business, soft-skill situational and procedural simulations as well. Such simulation vendors can be seen as the prototypes of MOOCs as we know them nowadays. [7]

2.3 Learning Pathways in MOOCs

Increased demand for online learning raised new challenges that concern fundamental aspects of learning. Dropout rates of 90% were commonly observed during a MOOC. [8] One of the main aspects of high dropout rates is low engagement.

Self-directed learning, which causes of low engagement in online education, has been one of the main modes of knowledge acquisition for many people since the beginnings of humanity. Self-directed learning is a study form which implies that learners are solely responsible for planning, implementing, and even evaluating their effort. This makes their learning behaviors and outcomes vary significantly.

The preset learning paths are able to solve the above-mentioned problems as their inter-curricular relationship in learning paths are predefined by the tutors. Taking the courses in a designed order leads to the smaller deviation within the learning outcomes. Most of the progressive xMOOCs mentioned in the first section, have introduced the learning paths to their learners, which indicates the use of the learning paths as the methodology in distance learning.

A good example is Coursera, which has adopted the methodology of learning paths. Each of the paths provides students with a complete guide to the highest growth careers. It usually includes a list of needed skills together with pre-requisites and a sequence of suggested courses that cover those. [9]

Figure 2. *Cousera's career learning paths, retrieved from https://blog.coursera.org/*

Each of the learning paths consists of recommended courses from different fields to provide student with valuable skills in the speciality area.

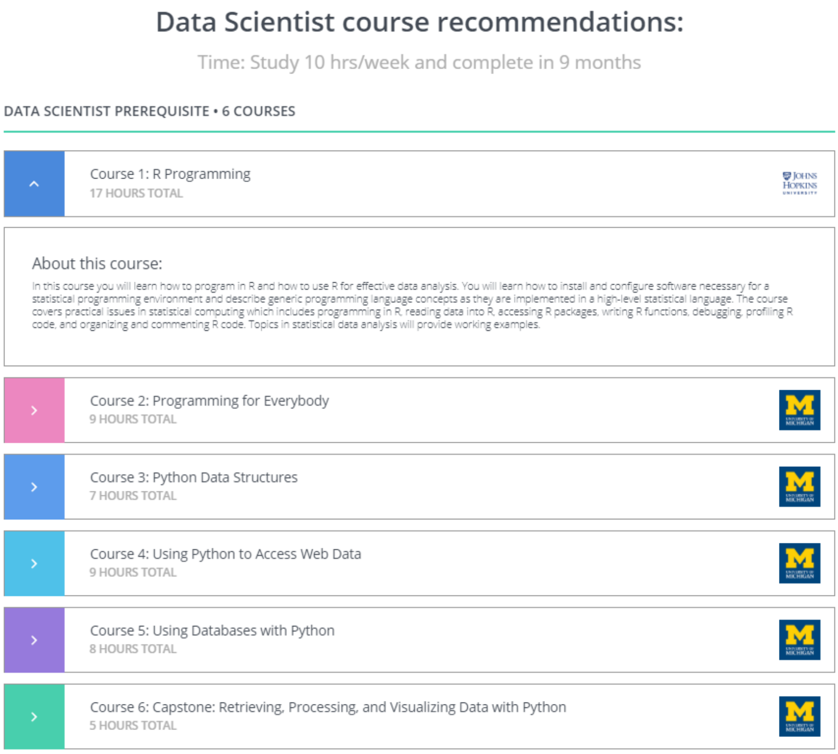

Figure 3. *Cousera's Data Scientist learning path, retrieved from https://careers.coursera.org/data-scientist/*

Similarly, other big players, such as, Khan Academy and Udemy have implemented the methodology of learning pathways for their own platform. Below, Table 2 visualizes some of the most successful xMOOCs and their implementation of the methodology of learning pathways.

Table 1. *A table of implemetation of the learning pathways in biggest xMOOCs*

xMOOC	**Learning Pathway**
Coursera	✓
Khan Academy	✓
Udemy	✓
edX	✓

Nowadays, the methodology of learning pathways has been adopted by the largest xMOOCs, which has proven to be successful over the past decade. Learning pathways help to direct learners to achieve their goals, and at the same time do not restrict them from taking the courses in preferred order. This makes learning pathways an essential part of the xMOOCs' strategy.

3. Structure of Learning Pathways

In this chapter the learning pathways in the scope of distance learning are being reviewed. Learning pathways are disassembled to get an understanding of what makes this pedagogical methodology have a positive impact on learning practice in xMOOCs. Next sections analyzes essential parts of learning pathways, which is later used for the design and analysis of learning pathways in the next chapter.

3.1 Designed v. Executed Learning Pathways

Before going into the principles of building the learning pathways it is important to draw a line between the two types of pathways. Davis, Chen, Hauff and Houben [10] took to attention how learners actually move through xMOOCs and called these movements **executed learning pathways**. In contrast, the learners' movements that the instructors/designers had in mind when creating the course were called **designed learning pathways**.

Figure 4 visualizes learning pathways taken in one week of a course in an xMOOC. On the left is what could be called a designed learning pathway of a classic xMOOC: first, the learner has to watch two videos, before answering 10+ quiz questions. Depicted on the right is how passing learners actually move through the course materials: check their scores, answer the quiz questions, avoid *binge watching* (watching one video after another). The difference between the visuals is the key argument to consider learning pathways in both designed and executed forms.

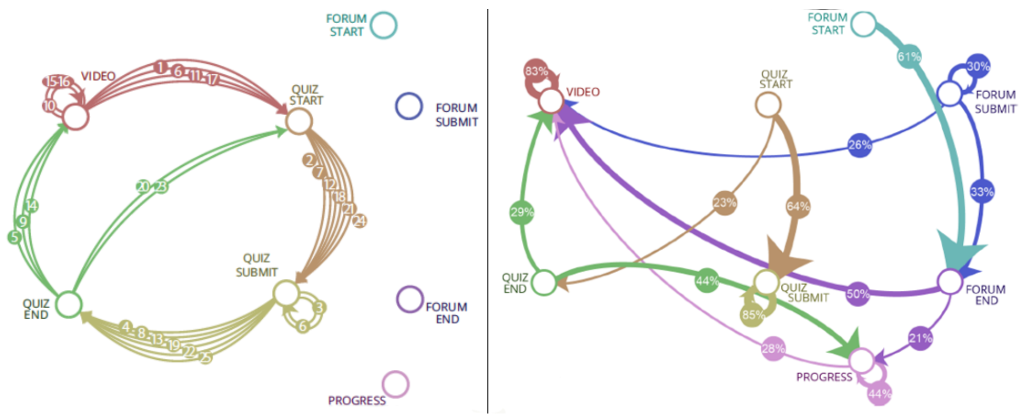

Figure 4. *Difference between designed (left) and executed (right) learning pathways, retrieved from https://chauff.github.io/*

3.2 Designed Learning Pathways

Designed learning pathways are linear learning sequences for building knowledge progressively from activity to activity and from week to week. Below are the main recommendations for building designed learning pathways.

3.2.1 Learning Outcomes

The key to creating a successful learning path strategy is to develop a plan that seamlessly blends the course essentials with the needs and wants of each learner. In addition to the individualized style and objectives they set for themselves, they must also complete their learning with a clear understanding of the key subject matter. Initially, Williams and Rosenbaum saw the end-goal in full *proficiency*, which meant that employees became independently productive. [2] xMOOCs set learning outcomes as a set of skills needed to become proficient in a certain field.

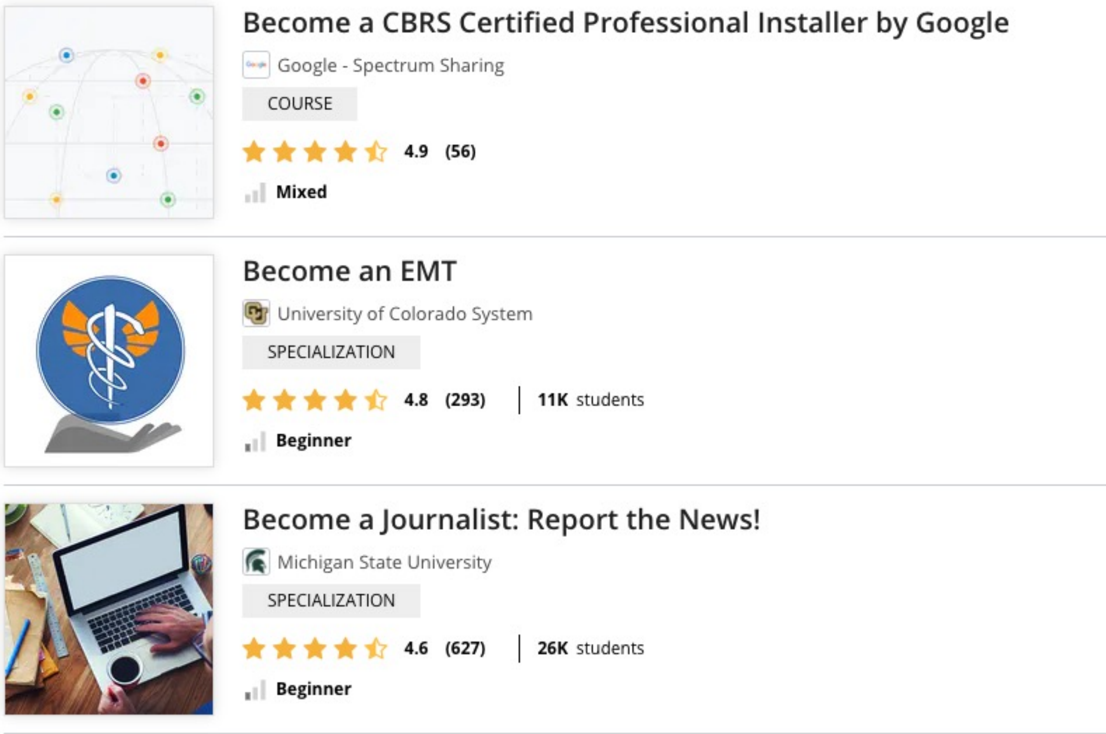

Figure 5. *Coursera's learning outcomes are represented as acquisition of career skills, retrieved from https://www.coursera.org/*

3.2.2 Strategic Prompts

A well-designed learning pathway does two things. It requires learners to activate their prior knowledge and to reflect on their initial experience after they have completed a certain module. A learning pathway designer needs to build-in prompts that focus on learners' experience and force them to apply their newly-acquired skills.

Figure 6 shows that based on the log traces, some passing learners jump backwards (from the middle to the beginning of Weeks 2 and 3). Backtracking can be used for reinforcement of prior knowledge to get over well-placed prompts.

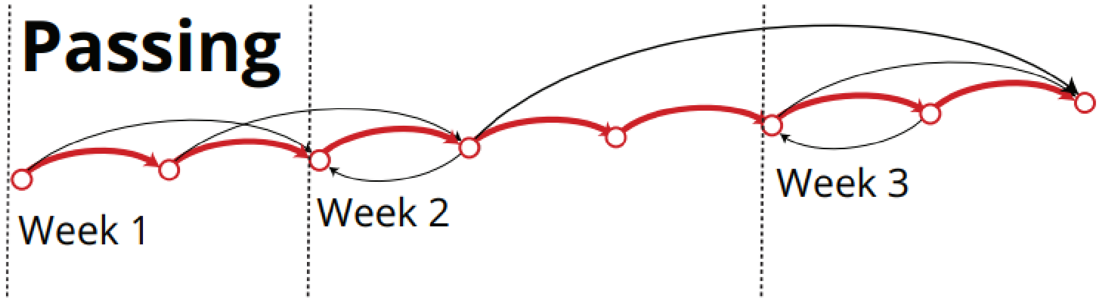

Figure 6. *Example of backtracking can be visible in the middle of Weeks 2 and 3, retrieved from http://www.educationaldatamining.org/EDM2016/*

3.2.3 Connections Between Stops

Educational approach and the buzzword of 2018 - *microlearning*, which offers bite-sized small learning units limited to the necessary amount of information. As the connections between different activities are not apparent, learning pathways' designers have to consider all of the stops made during a learning journey and make connections between them to bring the learners to their goals. Stops in place give learners a chance to check their progress along the way and ensure that they are on the right path.

Figure 7 gives visual representation of two possible stops' placements. The light-colored boxes identify learning activities while the dark ones represent assessment activities. Colors of the boxes represent different topics.

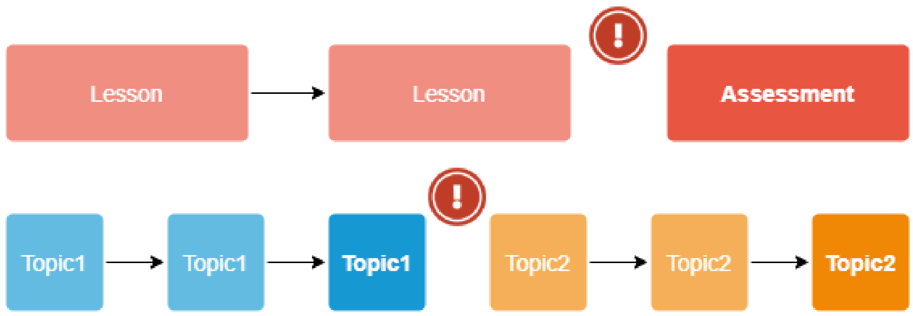

Figure 7. *Examples of Stops' Placements*

3.2.4 Scalability of Modules

As human knowledge extends over time and employees require more skills, xMOOCs should keep their educational materials up to date. That is why it is initially important to focus on systems that allow a flexible curriculum that can be updated or reassembled in different ways.

3.2.5 Visual Representation of Learning

Just like milestones represent a way of marking the way to the destination, a learning pathway needs posts to keep learners focused and on their target. This helps them see the progress they have made and what still needs to be done in order to complete a pathway.

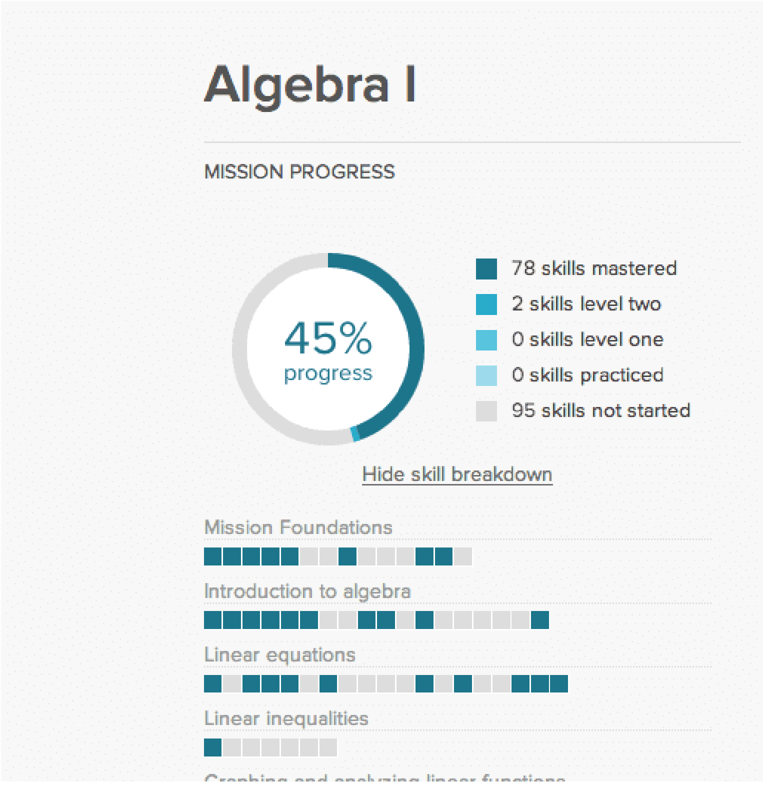

Figure 8. *Khan Academy's Dashboard, retrieved from https://www.khanacademy.org/*

3.3 Executed Learning Pathways

Tutoring systems with designed learning pathways have demonstrated an increase in effectiveness of learning. [11] However, even the most effective tutoring system fail if the students' behavior is not receptive to the material being presented. For example, lack of motivation has been shown empirically to correlate with a decrease in learning rate. In addition, students often use tutors ineffectively and adopt behavioral strategies that allow them to avoid learning, e.g., deliberately entering incorrect answers to elicit hints and, eventually, the correct answer from the tutor. Although tutor instruction is beneficial, its effectiveness might be increased if maladaptive student behaviors could be identified.

Executed learning pathways are the product of logged student actions during learning sessions. The actions are categorized, binned and symbolized. The resulting symbols are arranged sequentially, and examined by a motif discovery algorithm to detect repetitive patterns, or motifs, that describe frequent tutor events. These motifs are examined and categorized as student behavior. [12] The main goal of tracking student behaviour is detection of problem solving styles, either based on a description of pre-defined styles or with the aim to discover new ones. Additionally, Executed learning pathways are effective in predicting passing students, i.e., whether a student will successfully deal with a problem.

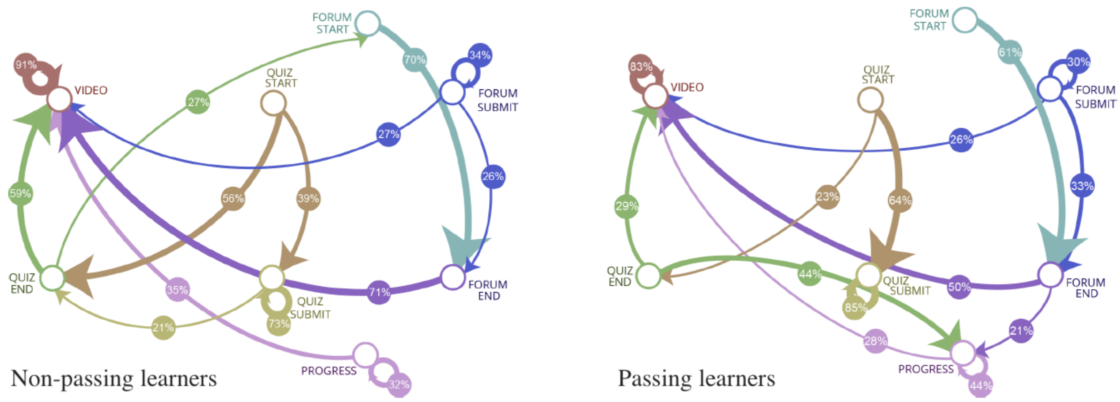

Figure 9. *Executed pathways of non-passing and passing learners, adapted from https://chauff.github.io/*

Below provided some of the key elements for building the executed learning pathways.

3.3.1 Tutoring System

Tutoring system directs students through the learning process. Functionality of tutoring systems may include:

1. Setting an objective or a problem to solve for each lesson.
2. Asking questions.
3. Providing immediate feedback.
4. Revealing hints, providing a problem's solution.
5. Grading student's performance during the lesson.

Tutoring systems and their functionality serve the needs of education vendors. While some teach students how to solve standardized tests (e.g. geometry, statistics and algebra problems), the others may provide space for practice in programming or simulate physical experiments.

3.3.2 Data Collection and Metrics

During interaction with a tutoring system, each student's actions are logged in a database. Metrics of log entries vary for the purposes of the research:

H. Chuang and C. Shen [13] analyzed the performance of two groups of students in Taiwan elementary school. Control group studied each unit at will, while Experimental group's learning path was controlled, students in this group had to learn in a fixed sequence. Metrics of this research were limited to the performance of groups on the final exam, see Figure 10.

Group	N	Pre-test	Post-test
		M	M
Experiment	140	30.12	72.12**
Control	144	30.23	64.07
**Learning effectiveness is significant at p<0.01			

Figure 10. *Metrics of H. Chuang and C. Shen's experiment, adapted from "A Study on the Applications of Learning Paths Conceptsto the Teaching in Elementary School", by H. Chuang and C. Shen, 2008.*

Another research, by Shanabrook, Cooper, Woolf and Arroyo [12], on Wayang Outpost – tutoring system that uses multimedia and animated adventures to help prepare middle and high-school students for standardized math tests.

M1	2305	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo
M2	6789	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo
M3	8989	adiq	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo
M4	8993	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo
M5	9485	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo
M6	2301	cdgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo	adgo
M7	18181	adgo	adgo	adgo	adgo	adgo	bdgo	adgo	adgo	adgo	adgo
M8	29469	afho	aeho	aeho	aeho	aeho	aeho	aeho	afho	aeho	aeho
M9	29301	aeho	aeho	aeho	aeho	aeho	aeho	aeho	afho	afho	aeho
M10	48953	aeho	afho	afho	afho	aeho	aeho	aeho	aeho	afho	aeho

Figure 11. *10 motifs from research on Wayang Outpost, adapted from "Identifying High-Level Student Behavior Using Sequence-based Motif Discovery", by by Shanabrook, Cooper, Woolf and Arroyo, 2010.*

Researchers utilized *four* metrics: hints seen (hints), seconds to first attempt (secFirst), seconds between subsequent attempts (secOther), and incorrect attempts (numIncorrect).

Once the metrics were binned, each problem was represented as a four character problem string, where each of characters represented a performance value from metrics. For any given student, the tutor interaction could be summarized by the sequence ordered concatenation of problem strings that researchers called a student string.

3.3.3 Patterns Discovery

Before the discovery of the patterns it is important to clear the dataset. Described by Chiu, Keogh, and Lonardi presence of noise in the dataset may limit the ability of researchers to discover motifs. [14]

Following this conclusion, Davis, Chen, Hauff and Houben removed all edges that have a probability below 20% to avoid clutter. [10] Other group of researches, Shanabrook, Cooper, Woolf and Arroyo [12], considered the repetition of undesired behavior as an important feature to capture, so the researchers did not remove such motifs as degenerate.

For patterns discovery researchers use mathematical systems and statistical hypotheses. Similarly to the metrics collection and filtering of the data noise, there is no single model or hypothesis. Each group of researchers uses different approach for the discovery of patterns. For example, Davis, Chen, Hauff and Houben utilized the *discrete-time Markov chains* in order to chart the likelihood that a learner will transition from one engagement activity to another. Figure 9 is the example of application of the discrete-time Markov chains on learner's motifs.

Previously mentioned researchers, H. Chuang and C. Shen [13] have built multiple statistical models and tools based on just one metrics:

1. **T-test** - *used to determine if there is a significant difference between the means of two groups.* Researcher compare the difference in pre- and post-test scores. The results indicate that the path-controlled group performed significantly better than the random group.
2. After discovering that path-controlled students performed better, the researches used AlterWind LogAnalyzer to identify the distribution of students' learning pathways. The distribution during the experiment can be seen in Figure 12

Learning paths	Total
Sequential	72
Iterative	34
Challenging	16
Free	18
Total	140

Figure 12. *Distribution of Learning pathways, adapted from "A Study on the Applications of Learning Paths Concepts to the Teaching in Elementary School", by H. Chuang and C. Shen, 2008.*

3. Applied **Analysis of variance (ANOVA)** model, which provides a statistical difference of means for more than two groups (while t-test only applies to the maximum of 2 groups). The result of ANOVA analysis can be seen in Figure 13

Groups	Means	Standard deviations
Sequential	66.20	1.07
Iterative	63.12	1.52
Challenging	63.22	3.54
Free	54.30	2.52

Figure 13. *Descriptive statistics of learning outcome in different learning path, adapted from "A Study on the Applications of Learning Paths Concepts to the Teaching in Elementary School", by H. Chuang and C. Shen, 2008.*

4. Analysis of Learning Behavior

This chapter describes an xMOOC with an absent learning sequence as a research subject of the experiment. The goals of the experiment are to retrieve, analyze, and justify the patterns of learners' behaviour. Hypothetically, this behaviour without learning sequence diverges significantly and obstructs the extraction of the patterns. The irrationality of learners' personal choices of accessing educational materials will be the root cause of the creation of designed learning pathways.

4.1 Research Subject

As a subject for development cyber security-aimed e-learning platform RangeForce (RF) has been picked. RangeForce is a fast-growing Estonian start-up, which provides educational modules to develop hand-on skills in cyber security. Educational material is provided in a form of virtual on-demand environments (modules) with different sets of problems and objectives. Modules are subjected to four topics: Web Application Security Essentials (WASE), Security Tools, Security Operations Center (SOC) and DevOps.

RangeForce modules also vary by type. Educational modules give students step-by-step guidance and objectives via a tutoring system called Virtual Teaching Assistant (VTA). In order to complete an objective, students are required to make changes to the machines in the virtual environment, answer multiple answers questions or submit the correct value into the prompt. VTA immediately provides feedback on the correctness of an answer. User can request hints and obtain a full solution of the objective from the VTA.

Aside from the educational modules, there are assessment modules of two types:

- **Challenges** - small CTF-like (Capture-The-Flag) modules, which are mostly based on material covered in educational materials. Usually, in challenges users are provided with no hints from the Teaching Assistant and are only given one objective - to submit the flag as a proof of reaching the target.
- **Sieges** - large scale assessment modules which have CTF-like scenarios and structure. Usually, they do not imply any hints or solution, or provide those for a set amount of points (points are acquired on the completion of each objective). Instead of a single objective, sieges combine multiple ones.

On the RangeForce learning platform (Userhub) learner gets a list of available modules on all kinds of topics without any learning sequence.

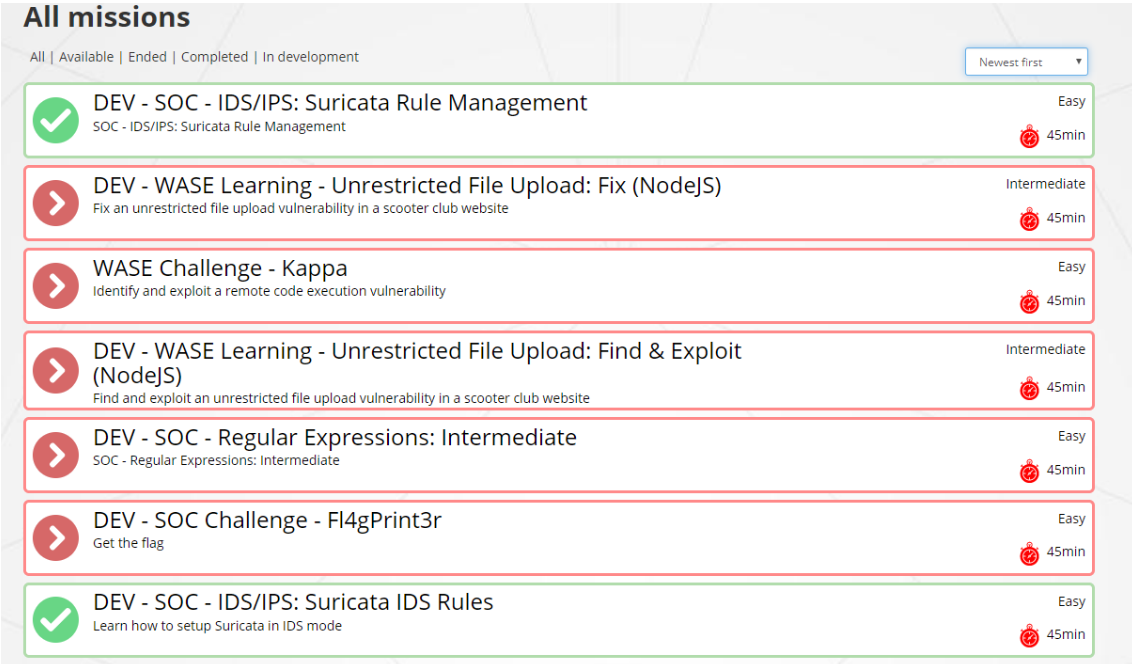

Figure 14. *View of RangeForce Userhub, retrieved from https://hub.rangeforce.com/learn*

In the context of this research the lack of order in RangeForce modules might lead to low engagement level of the learners which consequently will cause a high dropout rate. According to the analysis of the innovative xMOOCs in Chapter 2, the implementation of learning pathways can mitigate the negative consequences.

Tutoring System

Previously mentioned Virtual Teaching Assistant 1.4 (VTA) is the tutoring system of RangeForce. Most of its features are illustrated in Figure 17. It guides a learner during the module with the ordered list of steps. Once the learner begins the step, it provides a short story-based description and an objective to complete. It has different types of checks for evaluating the completion of the objective: prompts, multi-options questions, automated scripts for checking the correctness of the changes in the virtual environment. User can request hints and solutions for the objective. Also, it keeps track of the learners' progress in percentage and time.

Most importantly, students' interactions with the VTA are logged. Therefore, after collecting logs it is possible to analyze behavior and retrieve patterns of executed learning pathways.

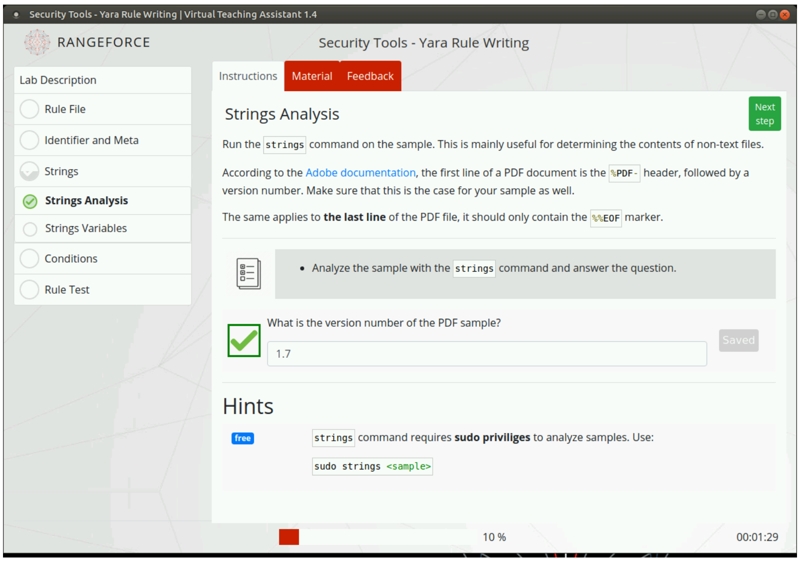

Figure 15. *Virtual Teaching Assistant 1.4 is the tutoring system of RangeForce, retrieved from https://hub.rangeforce.com/learn*

4.2 Analysis of Executed Learning Pathways

4.2.1 Ethical Considerations

Before building the metrics for the data analysis, it is important to address the ethical aspect of the research. All the data was collected in accordance with the RangeForce Privacy Policy, which states that the company has a right to collect and use 'performance data and information' about how one uses the RangeForce products and services.[15] The dataset contains the usernames, which may be used to directly or indirectly identify users. [16] Considering that personal data is not necessary for the data processing, usernames are *pseudonymized*. Pseudonymization is a de-identification procedure used to replace personally identifiable information fields within a data record with one or more artificial identifiers, or pseudonyms. The artificial replacement for the usernames has a structure of *User X*, where *X* is a numeric iterator. Below is an example of the usernames' pseudonymization in this research.

```
> users %>% select(username, X_id, role)
    username                        X_id    role
    User 8       56725c4127a51475d2000004   user
    User 13      568bbc3127a5143eea000000   user
    User 87      5698e78327a51436bc000006   user
    User 201     56c1913427a5142cdc000000   user
    User 294     5752b0676442ad6ff5481cee   user
```

Figure 16. *Pseudonymized usernames in users table, example in R language.*

4.2.2 Data Collection and Metrics

For data analysis in this research I have used R-language version 3.6.2 in R-studio environment version 1.2.5033 with dplyr package for manipulations with dataframes. The most important in the context of this research is the data table of the module attempts since December, 26, 2016. Module attempts are the logs of users' interactions with the Tutoring System. Initially, the data table contained 40217 log entries of 1988 variables.

Data Filtering

As some of the log entries in the dataset belong to the RangeForce employees (e.g. developers, testers). In order to avoid clutter in the analysis of learner's behaviour, all log entries with RangeForce-related identifiers were filtered out. As a result, I have obtained 14699 log entries created by the RangeForce learners.

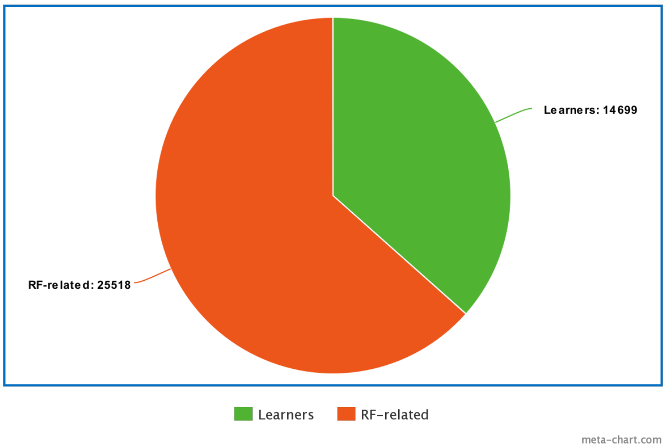

Figure 17. *Log entries by RF-related users and learners as of Februrary, 5, 2020*

With filtered and pseudonymized data I was able to start the analysis of the data. Firstly, I separated the individual log entries for each learner. Using iteration of the learners' identifiers, I have extracted timestamps of each attempt, identifiers of the modules, progress in percentage and the user id. As a result, I have obtained logs of modules attempts for 2781 users.

To not misinterpret the order of the modules taken, I considered only the attempts with progress higher than zero. Following this logic, *see Figure 18*, if *userID 2* had *a progress of 95.23810* percent in *moduleID 586b61aa6442ad3bf3ac8714* (line 1, Figure 18), I consider his next choice of the module valuable, which is *moduleID 57e391296442ad269b1509f4* (line 2, Figure 18).

```
> user_attempt <- filtered_lab_users %>% filter(labuser_id==2) %>%
                  select(started_at, lab_id, progress, labuser_id)
> show(user_attempt)
                 started_at                   lab_id progress labuser_id
1  2017-01-19T08:29:18.000Z 586b61aa6442ad3bf3ac8714 95.23810          2
2  2017-01-19T09:27:12.000Z 57e391296442ad269b1509f4  0.00000          2
3  2017-01-19T09:32:10.000Z 586b61aa6442ad3bf3ac8714 32.73810          2
4  2017-04-27T16:18:34.000Z 58b683b96442ade80d23980f  0.00000          2
5  2017-06-15T11:43:53.000Z 566ab45527a51401d3000007  0.00000          2
6  2017-06-16T12:23:34.000Z 566ab45527a51401d3000007  0.00000          2
7  2017-06-21T11:12:19.000Z 57a487616442ad320f21ec93  0.00000          2
8  2018-02-28T10:20:31.000Z 566ab45527a51401d3000009  0.00000          2
9  2018-04-09T10:06:34.000Z 5964a135d1069292d045f2df 37.03704          2
10 2018-07-26T12:04:44.000Z 5b3c854aa0880370f8d3840f  0.00000          2
11 2018-09-21T14:46:25.000Z 566ab45527a51401d3000009  0.00000          2
12 2018-12-04T09:46:49.000Z 566ab45527a51401d3000009  0.00000          2
13 2019-02-25T07:47:29.000Z 566ab45527a51401d3000009 38.09524          2
14 2019-02-28T22:53:55.000Z 5c41d767a08803351bf36a7e  0.00000          2
15 2019-04-02T15:38:38.000Z 5c41d767a08803351bf36a7e  0.00000          2
16 2019-04-17T11:39:30.000Z 566ab45527a51401d3000009 95.23810          2
17 2019-06-06T13:35:34.000Z 57a487616442ad320f21ec93  0.00000          2
18 2019-09-04T05:48:55.000Z 5c41d767a08803351bf36a7e  0.00000          2
```

Figure 18. *All extracted module attempts of UserID 2, example in R language.*

With the same logic in mind, I have collected the relations between modules for every user's attempts, which could be interpreted as *"after completion of the **lab1** with **progress X** user took **lab2**"*.

```
                        lab1         progress                     lab2
1   57a487616442ad320f21ec93              100 5c41d767a08803351bf36a7e
2   566ab45527a51401d3000007 6.34920634920635 59e9ec67d10692a8f0ceb8e9
3   586b61aa6442ad3bf3ac8714 95.2380952380952 57e391296442ad269b1509f4
4   586b61aa6442ad3bf3ac8714 32.7380952380952 58b683b96442ade80d23980f
5   5964a135d1069292d045f2df  37.037037037037 5b3c854aa0880370f8d3840f
6   566ab45527a51401d3000009 38.0952380952381 5c41d767a08803351bf36a7e
7   566ab45527a51401d3000009 95.2380952380952 57a487616442ad320f21ec93
8   57e391296442ad269b1509f4               75 58b683b96442ade80d23980f
9   566ab45527a51401d3000009 17.3160173160173 566ab45527a51401d3000009
10  5964a135d1069292d045f2df  37.037037037037 5b3c854aa0880370f8d3840f
```

Figure 19. *Representation of users' next steps after completing modules with the progress higher than zero, example in R language.*

The data table of relations between modules contains 2182 entries. Charting of 145 vertices and 2182 edges contains a lot of clutter, so I had to search for the strongest patterns. Based on the research of Chiu, Keogh, and Lonardi and the example of Davis, Chen, Hauff and Houben, I have removed edges with weak relations and only left the strongest ones. [10] [14] To find the strongest relations between the labs, I have iterated through the list of modules and saved the entries with the highest progress value for plotting. As earlier lab entries were identified by hashes I reassigned the names to each lab.

```
for (ent in 1:length(unique(valuable_labs$vertice1))){
    top <- unique(valuable_labs
            %>% filter(vertice1 == unique(valuable_labs$vertice1)[ent])
            %>% top_n(10, as.double(weight)))
    for(i in 1:length(top$vertice1)){
        if (top$vertice1[i] == top$vertice2[i]){ next()
        } else {
        rename1 <- as.data.frame(labs %>% filter(X_id == top$vertice1[i])
                    %>% select(name))
        top$vertice1[i] <- as.character(rename1$name)
        rename2 <- as.data.frame(labs %>% filter(X_id == top$vertice2[i])
                    %>% select(name))
        top$vertice2[i] <- as.character(rename2$name)
        }
    }
}
```

Figure 20. *Extraction of vertices with the highest progress, example in R language.*

4.2.3 Visualization and Patterns Discovery

I have separated the modules by topics (DevOps, Security Tools, SOC and WASE) to visualize the inner relations between modules of the same topic and relations to the modules of different topics. For visualization and plotting I have used the igraph package for R-studio, the snippet of visualization script presented in Figure 21.

```
library(igraph)

g_topic <- graph_from_data_frame(dataframe, directed=TRUE)
tkplot(g_devops,
       canvas.width = 1900,
       canvas.height = 1000,
       layout=layout_on_grid(g,
                             width = 5,
                             height = 5,
                             dim = 2) )
```

Figure 21. *Script used for visualization of the dataframe, example in R language.*

DevOps Topic

Figure 22 visualizes the inner and outer relations of modules which originate from the DevOps topic. As it can be seen, there are only a few strong inner relations in the DevOps topic: between vertices DevOps - IDS/IPS Security, DevOps - Introductory Lab, and DevOps - User Management. All of them are bilateral, which means multiple learning patterns are present between modules hence the order in which they were passed cannot be retrieved. This bilateral edges demonstrate that without any learning sequence, such as a learning pathway, self-paced learning might be irrational.

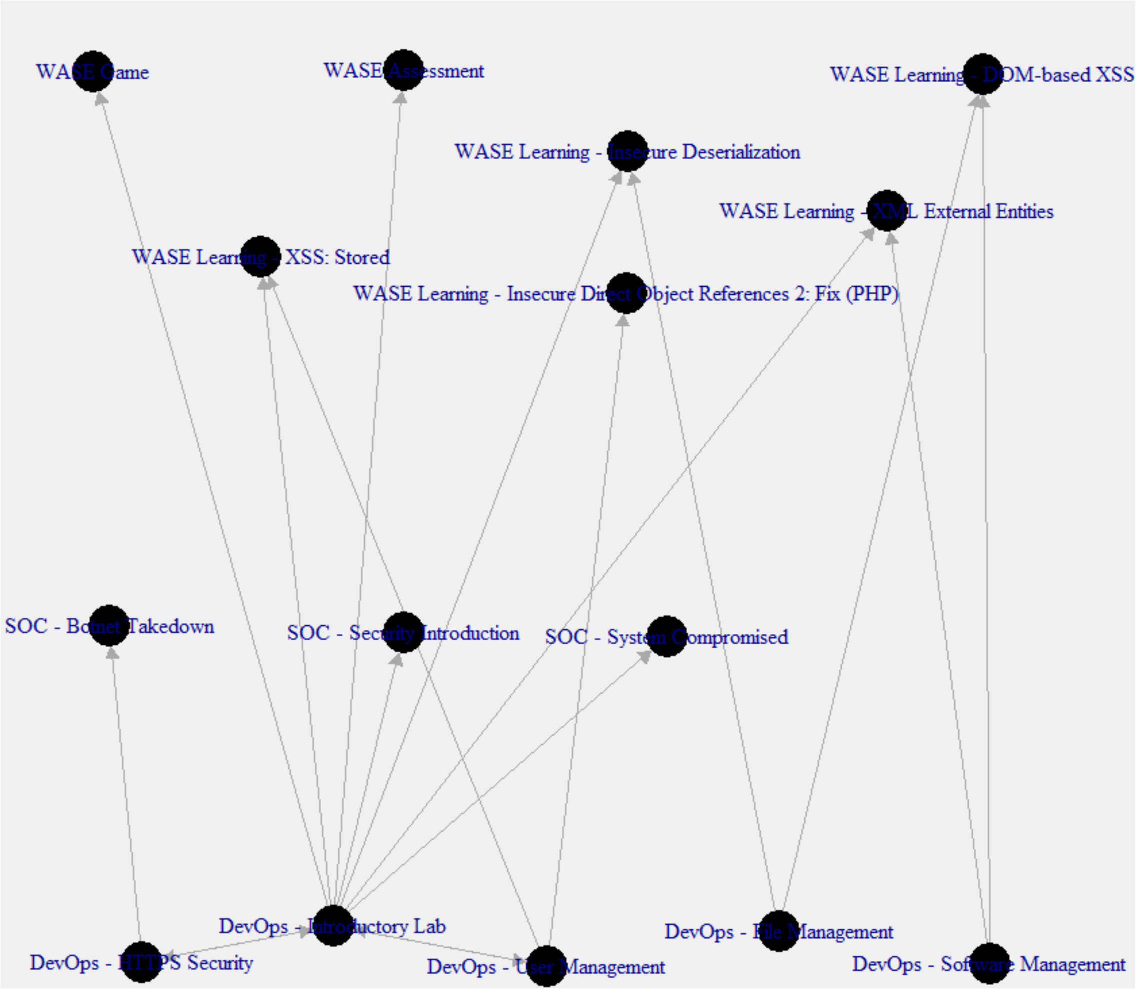

Figure 22. *Visualization of module relations which take origin from DevOps topic.*

Some patterns could be also retrieved from out of the topic relations. Using *DevOps - Introductory Lab* as an example, it can be seen that the module has relations with seven modules of different topics. This patterns of relations can be explained by the early marketing strategy of RangeForce. In the past years, RangeForce provided batches of modules via promo codes to the prospect customers. The *DevOps - Introductory Lab* was included in 199 promo codes. The distribution of the promo codes strengthened the relations between modules in the same batch. Figure 23 illustrates multiple promo codes which involve *DevOps - Introductory Lab*, for privacy reasons some of the labels are blurred.

Labels
[redacted], intro=∞
intro=∞, [redacted]
bot-takedown=∞, devops-user=∞, https-sec=∞, intro=∞, [redacted], secintro=∞, [redacted], wase1-sqli=∞
bot-takedown=∞, devops-user=∞, [redacted], intro=∞, secintro=∞, wase1-path=∞, wase1-sqli=∞
intro=∞, secintro=∞, wase1-sqli=∞
intro=∞, wase1-sqli=∞

Figure 23. *Examples of promo codes which involve DevOps - Introductory Lab.*

Web Application Security Essentials Topic

The modules of the WASE topic illustrate the reverse pattern of relations to the previously mentioned *DevOps - Introductory Lab*. While internally the modules belong to the same sub-groups, for example, *WASE Learning - XML External Enities* and *WASE Learning - XML External Enities 2*, *WASE Learning - XSS: Reflected* and *WASE Learning - XSS: Stored* and *WASE Learning - XSS: Stored-based Phishing*, they show weak or no relations at all. Based on this evidence, I assume that absence of the learning sequence makes learners to complete the modules in an order which does not allow to reveal a learning pattern.

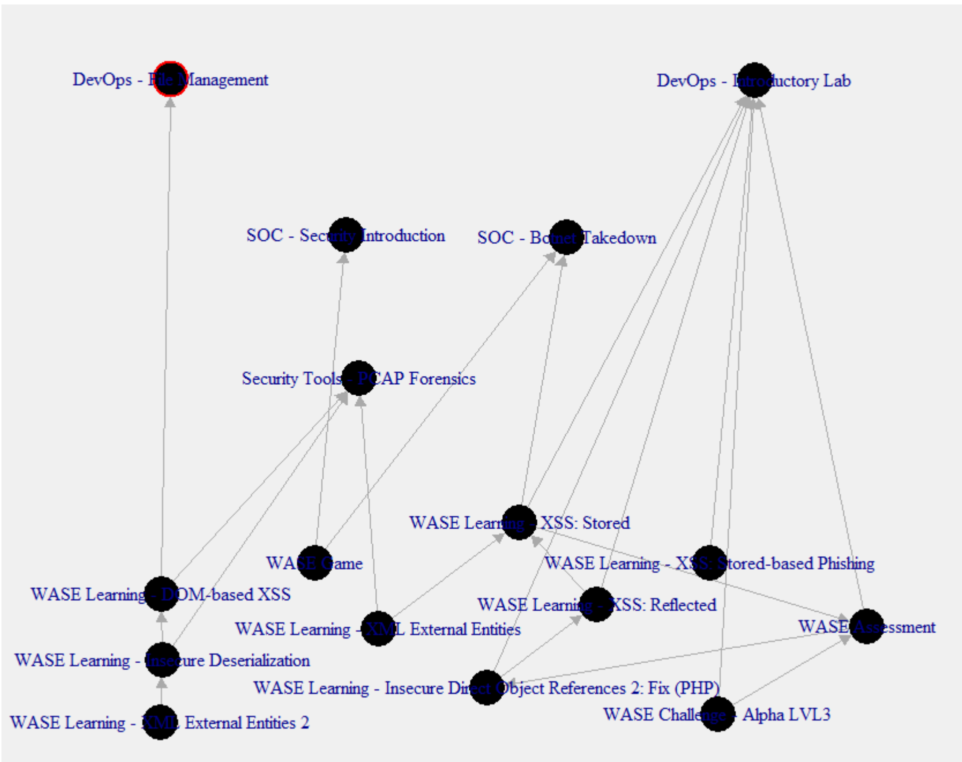

Figure 24. *Visualization of module relations which take origin from the WASE topic.*

Security Operations Center Topic

Similarly to the previous topics, relations in the SOC topic are based on the batches of the promo codes. Special interest presents external relations with the event-based modules. Such as the Cyber Security Challenge UK (CSCUK17 and CSCUK18), Microsoft (MS and M365), and Siege of Utopia on-sight events. As these modules were specifically developed for these events, they require presence of the RangeForce employees, hence the modules are not available for the learners in the Userhub on-demand. Strong relations of the SOC topic with the on-sight events suggest that as of February 5, 2020, RangeForce has not collected enough module attempts from the Userhub to balance out the perfomance of the participants from the on-sight events.

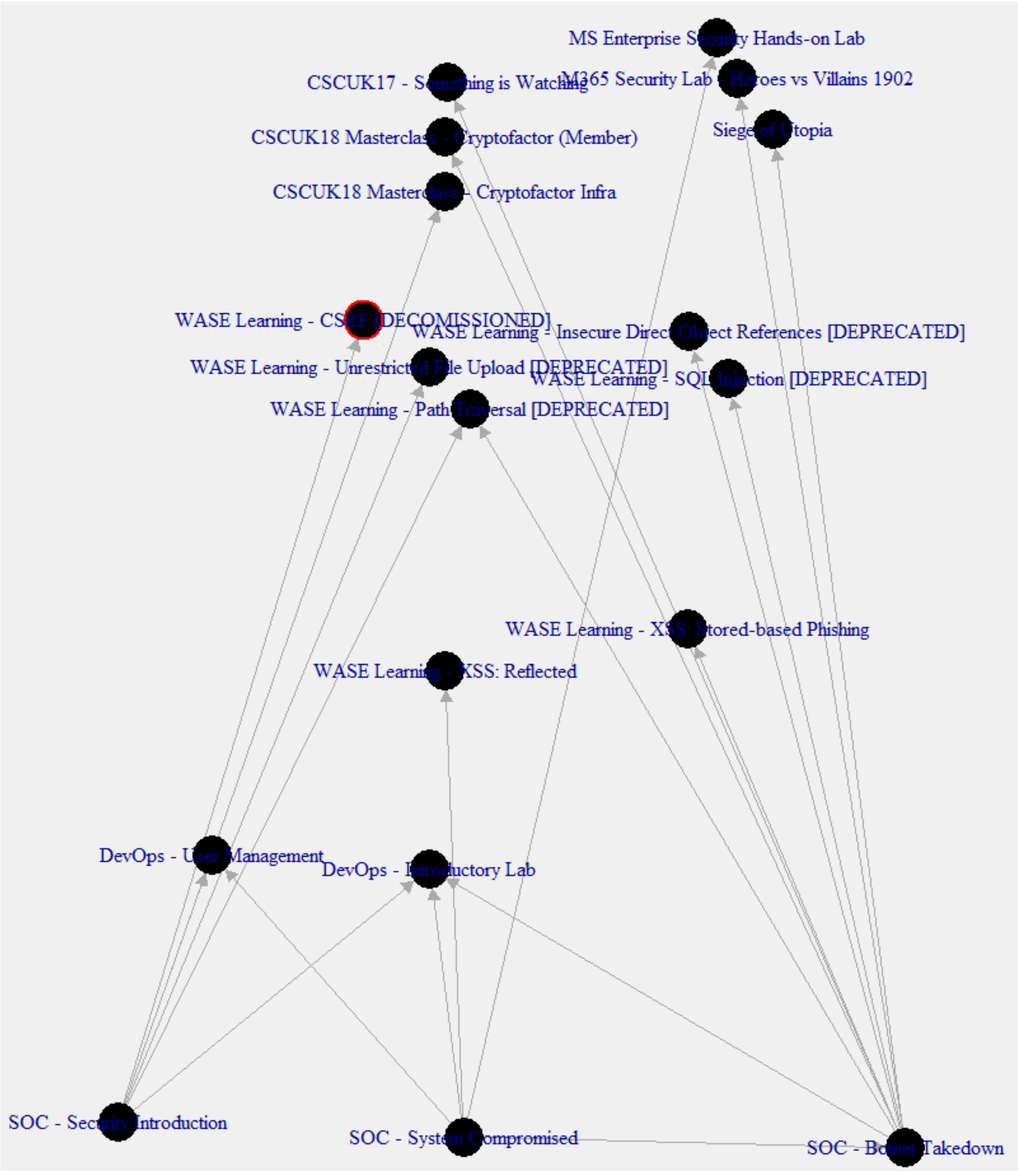

Figure 25. *Visualization of module relations which take origin from the SOC topic.*

Security Tools Topic

As it can be seen from Figures 22, 24, and 25, no modules lead to the topic of Security Tools. Absence of internal relations between the modules of the same topic was caused by the fact that learners did not pass the modules of the same topic one by one, or did not perform well during the module.

The explanation of this pattern could lie in the list of RangeForce modules. *Security Tools - Malware Analysis* and *Security Tools - PCAP Forensics* are listed as the 49th and 50th developed by RangeForce modules, which used to be a part of another topic - DevSecOps. While promo codes with modules of other topics were distributed to the customers, Security Tools (DevSecOps) modules were only accessed through promo codes twice, both times internally.

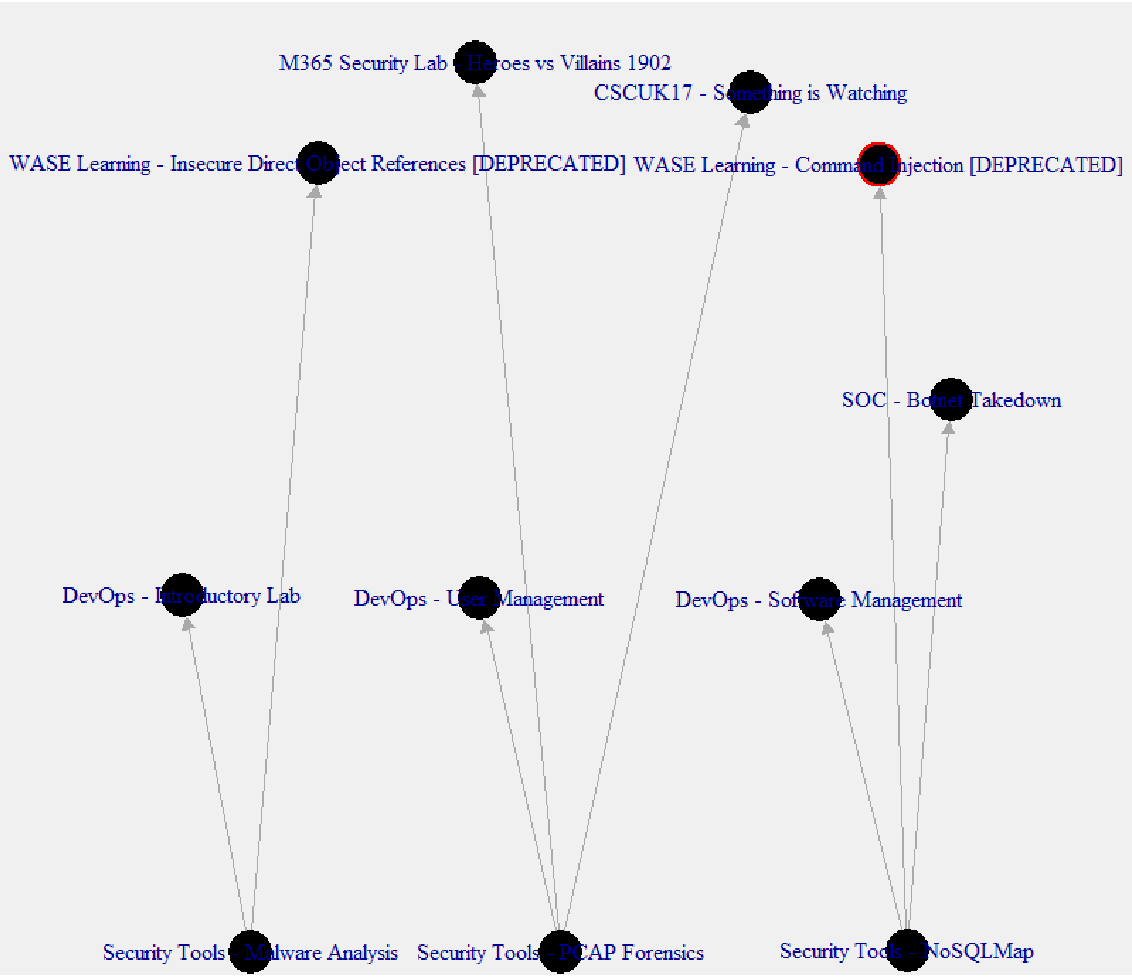

Figure 26. *Visualization of module relations which take origin from Security Tools topic.*

4.2.4 Outtakes

As the obtained patterns visualize the early strategy of the distribution of the RangeForce modules, the absence of a designed learning pathway made learners take modules based on their personal preference. As of February 5, 2020, the amount of modules completed by learners is too small to retrieve any learning patterns, see Figure 27. Top 10 learners have completed from 12.4 to 7.5 percent of learning material.

2	18
3	17
4	14
14	13
5	13
6	13
7	13
1	11
15	11
18	11

Figure 27. *Top 10 users by the number of module attempts, left column - pseudonymized id, right column - number of module attempts.*

Another evidence of irrationality of users' choices is the uneven distribution of the their attempts in each module. As illustrated in Figure 28, more than half of the modules have less than 100 attempts, and only a few have one thousand and more attempts.

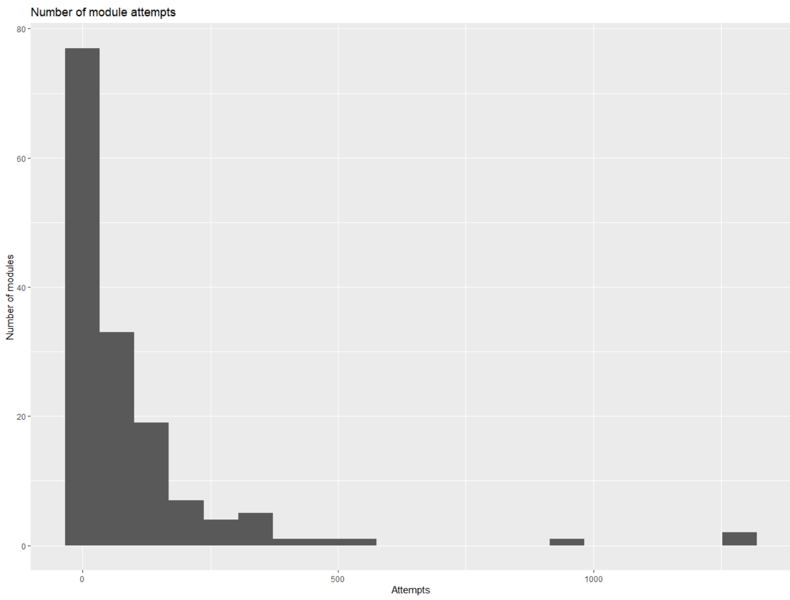

Figure 28. *Top 10 modules by the number of attempts, left column - module id, right column - number of module attempts.*

Taking this into consideration we can conclude that there are patterns of low engagement and disorientation. As was mentioned earlier, these negative consequences could be mitigated by developing designed learning pathways.

5. Development of Designed Learning Pathways

5.1 Development of Designed Learning Pathway

For the research purposes I have used an opportunity to analyze the dump of the RangeForce's databases. As of February 5, 2020, RangeForce had 145 modules listed. Full list of the modules can be found in Appendices. As some of the modules by the moment of analysis were not accessible from the Userhub, I manually filtered out all deprecated, Microsoft-based, and developed for on-sight events modules, - as they are not available for regular learners hence are not valuable for the development of learning pathways.

Table 2. *Table of filtered out modules by category*

On-sight events	**Depreated**	**Microsoft-related**
CSCUK17 - Something is Calling Me	WASE Learning - CSRF [DECOMISSIONED]	M365 Security Lab - Heroes vs Villains
CSCUK17 - Something is Watching	WASE Learning - CSRF Defend [DECOMISSIONED]	M365 Security Lab - Heroes vs Villains 1902
CSCUK17 Face2Face - Something is Wrong (Leader)	WASE Learning - Insecure Direct Object References [DEPRECATED]	MS - Microsoft Security Defender ATP
CSCUK17 Face2Face - Something is Wrong (Member)	WASE Learning - Path Traversal [DEPRECATED]	MS Enterprise Security Hands-on Lab
CSCUK18 Masterclass - Cryptofactor (Member)	WASE Learning - SQL Injection [DEPRECATED]	MS Enterprise Security Hands-on Lab - Walkthrough
CSCUK18 Masterclass - Cryptofactor Infra	WASE Learning - Unrestricted File Upload [DEPRECATED]	MS Ready Hacker Lab

Continues...

Table 2 – *Continues...*

On-sight events	**Depreated**	**Microsoft-related**
CSCUK19 - Qualifier 2	WASE Learning - Cookie Security: Secure [DEPRECATED]	Windows Defender ATP Workshop
CSCUK19 Face2Face - API Bleed (Member)	WASE Learning - Cookie Security: HttpOnly [DEPRECATED]	
CSCUK19 Face2Face - API Bleed Infra	WASE Learning - Command Injection [DEPRECATED]	
Hackerloo	DevSecOps - IDS/IPS Lab 1	
Platform Test Lab	Vulnerable Web Applications	
Siege of Utopia	WASE Game	
Trapped		

5.1.1 Connections Between Stops

As a result of filtering, I have obtained 111 modules which are available to the learners on RangeForce. I have manually completed each modules, and later analyzed their complexity and difficulty. As a result of the analysis, I have organized the modules in sub-groups (e.g Docker and HTTPS are the sub-groups of the DevOps topic) by the technologies or techniques, and connected the sub-groups by difficulty for the beginners. This way the learner starts with the sub-group of the introductory modules, and connections bring him to more complex sub-groups, which explain new tools and technologies thus reinforce the knowledge. Initial model of my analysis included logical conjuction and disjunction if sub-groups or modules had similar complexity and the order of passing these modules did not matter for the learning outcomes. However, naturally the logical disjunction means 'do X or Y, but not both' so it excludes some of the modules. In order to simplify the models, I have transitioned them to the linear ones. Linear models incorporate all the modules in a strict order, and consequently are more effective for the development of learning pathways. Figure 29 (on the left) represents the connections between the sub-groups of the DevOps topic.

Logical **ANDs** and **ORs** make the learning pathways complex and do not necessary include all the modules of the topics.

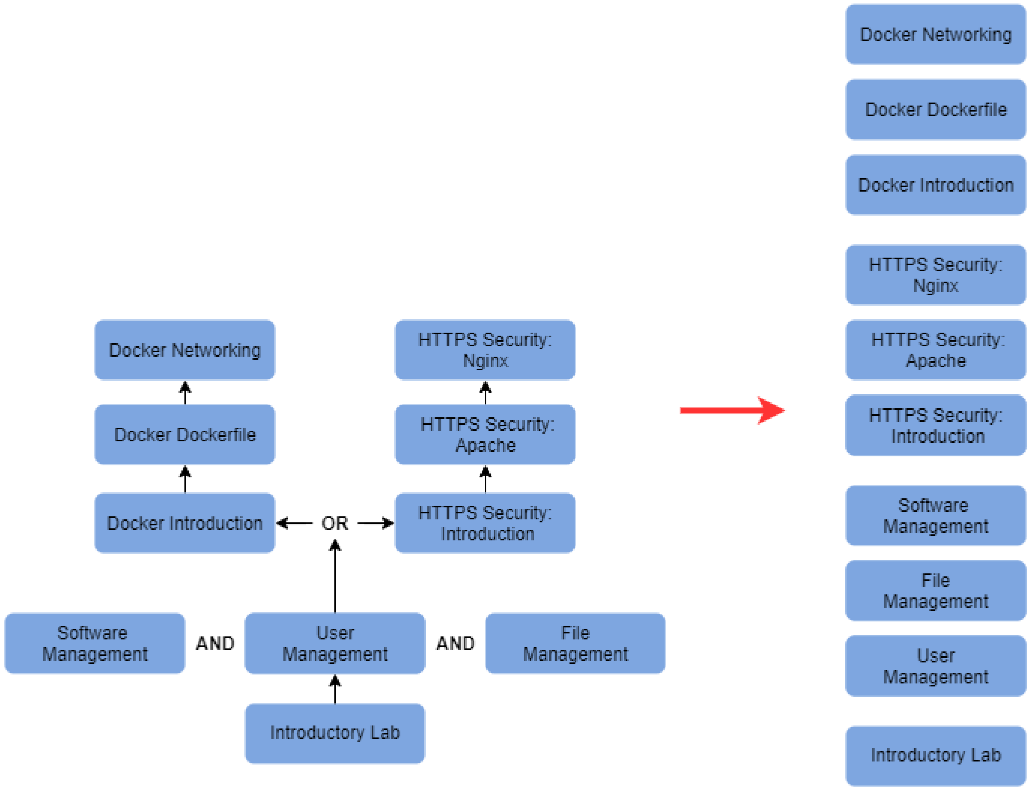

Figure 29. *Transition of the initial model of DevOps topic to the linear model*

I repeated the same procedure for every topic and have obtained the linear models of all available modules separated by the topics. The models in Figure 30 visualize the recommended order of learning each topic. Creation of the ordered list of modules by priority is necessary for building designed learning pathway. The only anomaly can be seen in the WASE topic (green in Figure 30), where modules with the same structure have multiple implementations in different programming languages (PHP and NodeJS). As most of the modules are implemented on the PHP programming language, it has been decided that the implemented on NodeJS modules should be left optional.

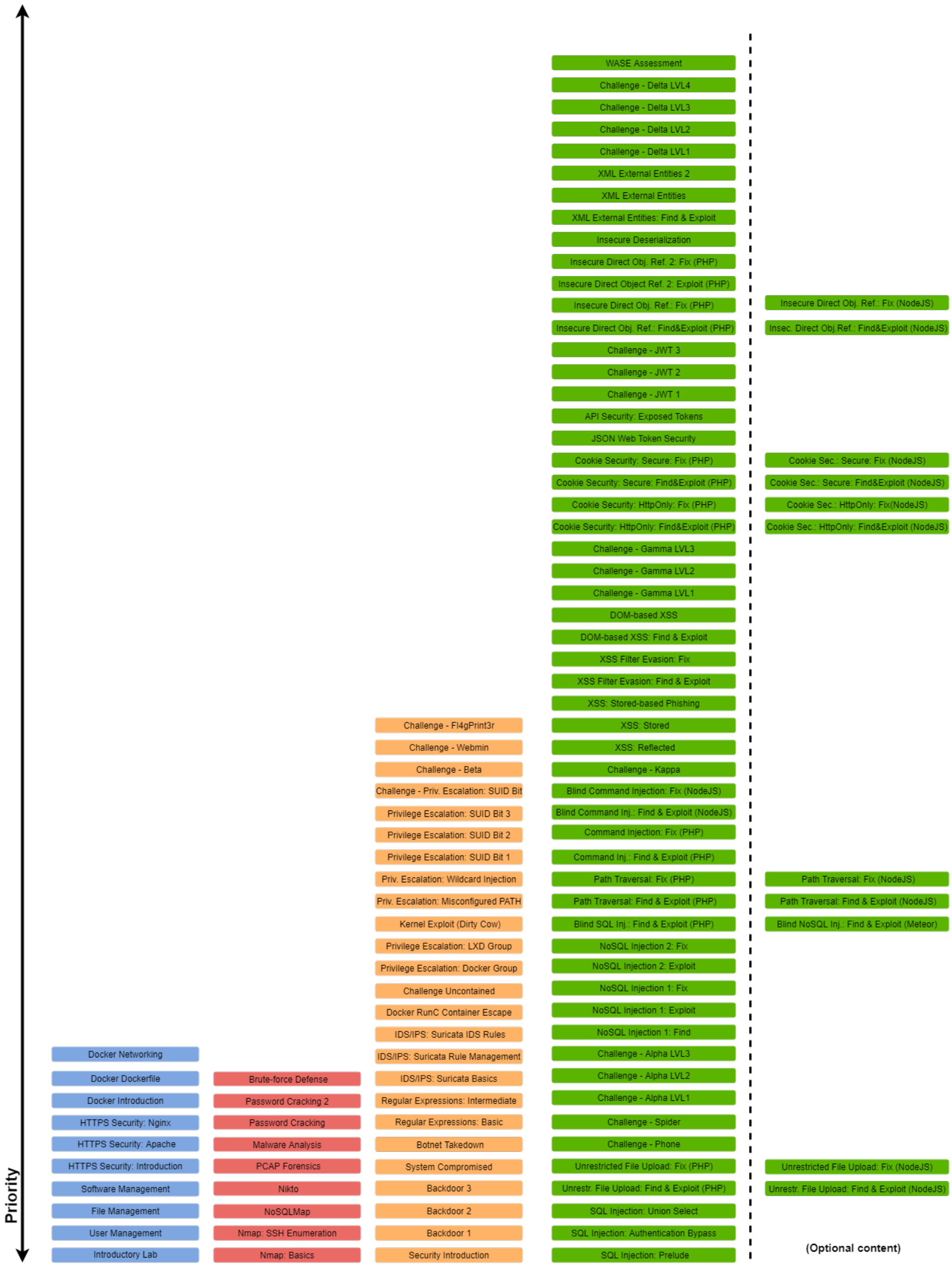

Figure 30. *Ordered lists of topics: blue - DevOps, red - Security Tools, orange - SOC, green - WASE*

5.1.2 Learning Outcomes

Despite every RangeForce modules has its own learning outcomes, it is important to set general learning outcomes which are achieved after the completion of a full learning pathway. It is possible to set generalized outcome as *'learning the DevOps topic'*. But, it is been discovered in Section 3.2.1 that large xMOOCs set the learning outcomes as becoming proficient in a certain field.

To comply with the modern standards, RangeForce has mappings of content in compliance with the National Institute of Standards and Technology[1] National Initiative for Cybersecurity Education[2] (NIST-NICE) Cybersecurity Workforce Framework and The MITRE Corporation's[3] Adversarial Tactics, Techniques & Common Knowledge[4] base. Both of these frameworks serve as fundamental reference points for describing and sharing information, the point is in bringing communities together to develop more effective cyber security. In the context of this research I will use the advantage of the NIST-NICE framework, as it is improves communication about how to identify, recruit, develop, and retain cyber security talent. [17]

Pooling RangeForce modules under the specialty areas of NIST-NICE framework will benefit in building the learning pathways for those specialties and will specify learning outcomes following the example of the leading xMOOCs. I have researched the NIST-NICE framework Specialty Areas and have analyzed the RangeForce modules and created a table of correspondence for each of 111 modules to 33 Specialty Areas descriptions. The example of the table of correspondence is presented in Figure 31, where matching modules are aligned with the Specialty Areas' descriptions.

		Securely Provision			
Lab	Production	Risk Management (RSK)	Software Development (DEV)	Systems Architecture (ARC) / Security Architect	Technology R&D (TRD)
Wase Learning - XSS Filter Evasion: Fix	done		x		x
Wase Learning - Cookie Security: HttpOnly: Find & Exploit (NodeJS)	done				x
Wase Challenge - Delta LVL4	done				x
SOC Challenge - Privilege Escalation - SUID Bit	done			x	

Figure 31. *Mapping of RangeForce modules to NIST-NICE's Specialty Areas*

[1]National Institute of Standards https://www.nist.gov/

[2]National Initiative for Cybersecurity Education https://www.nist.gov/itl/applied-cybersecurity/nice

[3]The MITRE Corporation https://www.mitre.org/

[4]Adversarial Tactics, Techniques & Common Knowledge https://attack.mitre.org/

Based on the mapping represented in Figure 31 I have designed learning pathways according to the complexity order represented in Figure 30. Below, Table 3 explains the NIST-NICE Specialty Areas.

Table 3. *Descriptions of NIST-NICE Specialty Areas: Incident Response (CIR) and Exploitation Analysis (EXP)*

Category	**Specialty Area**	**Description**
Protect and Defend (PR)	Incident Response (CIR)	Responds to crises or urgent situations within the pertinent domain to mitigate immediate and potential threats. Uses mitigation, preparedness, and response and recovery approaches, as needed, to maximize survival of life, preservation of property, and information security. Investigates and analyzes all relevant response activities.
Analyze (AN)	Exploitation Analysis (EXP)	Analyzes collected information to identify vulnerabilities and potential for exploitation.

In Figure 32 presented are the examples of two learning pathways for the Incident Response (CIR) and Exploitation Analysis (EXP) Specialty Areas, which are obtained from matching modules to the NIST-NICE descriptions and ordered accordingly to the complexity list from Figure 30. Learning pathways combine modules of different topics to reinforce the knowledge necessary for the Specialty Area.

As the RangeForce modules are independent entities, they sustain the concept of microlearning, so learners can make stops after completing every module. On the other hand, stops can be made after finishing one of the topics in the learning pathways - those are represented by different colors. For the learning pathways based on the NIST-NICE Specialty Areas, general learning outcomes can be set easily and clearly. For the pathways below suggested learning outcomes are "Become an Incident Responder" and "Become an Exploitation Analyst".

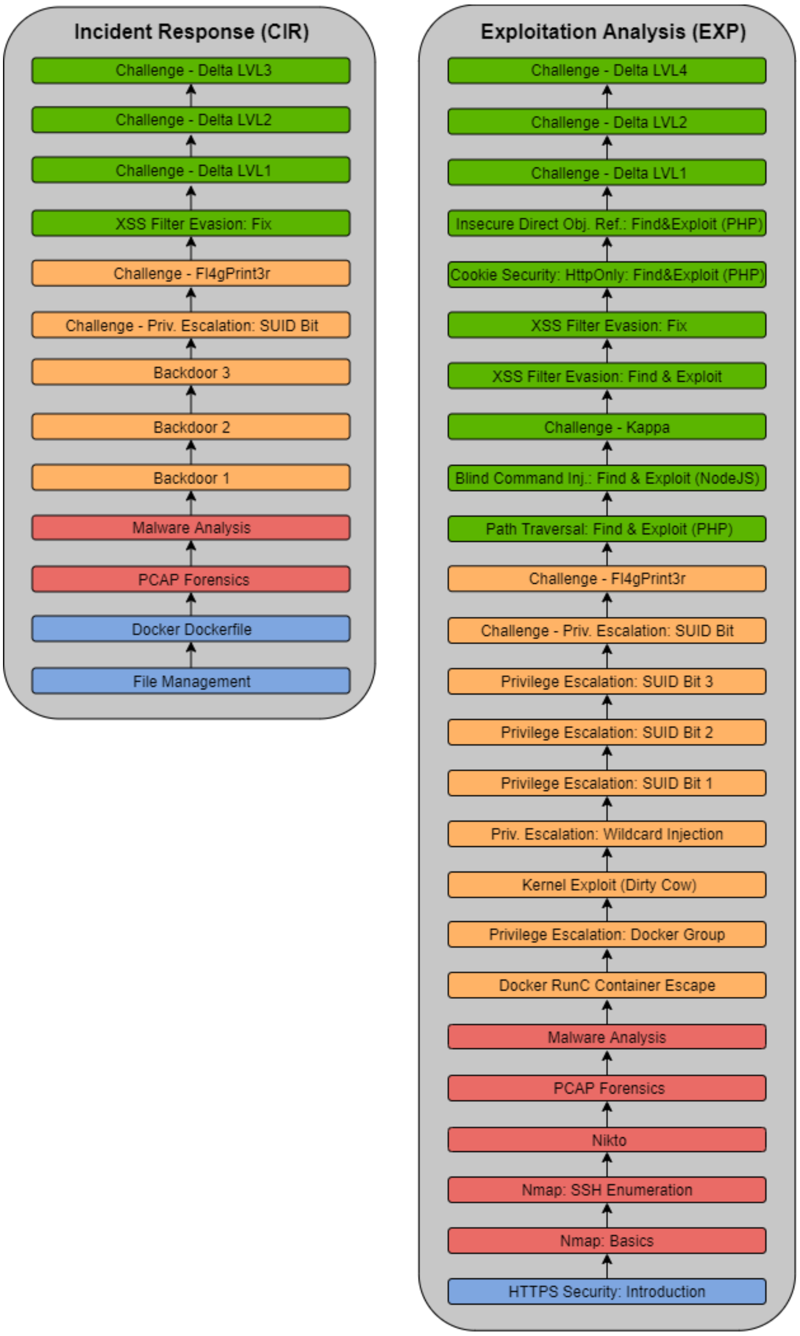

Figure 32. *Incident Response (CIR) and Exploitation Analysis (EXP) learning pathways*

5.1.3 Scalability of Modules

The RangeForce modules are independent entities, and company targets to release up to 20 modules monthly. In order to update the learning pathways, the similar process of evaluation of the complexity of the module and matching the content to the NIST-NICE Specialty Areas descriptions has to be carried out beforehand.

6. Summary

After gaining significant popularity, MOOCs discovered issues of learners' disorientation and low engagement. Over the time, learners' disorientation caused deviations from the learning objectives while low engagement increased the dropout rate. To mitigate this problems, designers of the xMOOCs adopted the concept of learning pathways. Initially developed to make orientation programs in large companies easier, learning pathways became an essential part of the most innovative xMOOCs. Aside from the designed learning pathways for guiding learners through the process of education, executed learning pathways help analyse personal learning style and performance. This paper reviews the cases of application of both types of learning pathways and describes their main aspects.

For the analysis of learners' behavior, data on the module attempts was collected and filtered. I have applied analysis of the modules of four separate topics and attempted to establish to what extent learners' behaviours are rational when there is no learning path to follow. During the analysis it has been discovered that due the lack of designed learning pathways in place and RangeForce's concentration on the on-sight events, the patterns of low engagement, high behavioral diversity, and disorientation were present.

Specifically, this work applies collected knowledge for developing learning pathways for security-aimed xMOOC RangeForce. As the learning pathways were absent on the RangeForce's platform, common for xMOOCs problems could appear in the future. To avoid these problems, based on the analysis of the complexity of the modules and application of the NIST-NICE framework Specialty Areas descriptions, the designed learning pathways were developed. Two examples of designed learning pathways which match Incident Response (CIR) and Exploitation Analysis (EXP) NIST-NICE Specialty Areas were introduced. Examples of learning pathways set the clear learning outcomes which match the highest growth careers in the field of cyber security.

Bibliography

[1] C. Peng. "The design and implementation of learning path management module of online education system". In: *The 27th Chinese Control and Decision Conference (2015 CCDC)*. 2015, pp. 5711–5714.

[2] Steve Williams Jim; Rosenbaum. "Learning Paths Increase Profits by Reducing the Time it Takes Employees to Get Up-To-Speed." In: *Learning Paths Increase Profits by Reducing the Time it Takes Employees to Get Up-To-Speed.* 2004.

[3] Laura Pappano. *The Year of the MOOC*. [Accessed: 15-04-2020]. URL: https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html.

[4] Alan Rothman. *A Real Class Act: Massive Open Online Courses (MOOCs) are Changing the Learning Process*. [Accessed: 20-04-2020]. URL: https://subwayfold.com/tag/mooc-hype-fades/.

[5] Troy Onink. *Georgia Tech, Udacity Shock Higher Ed With $7,000 Degree*. [Accessed: 20-04-2020]. URL: https://www.forbes.com/sites/troyonink/2013/05/15/georgia-tech-udacity-shock-higher-ed-with-7000-degree/.

[6] FutureLearn. *FutureLearn delivers the largest MOOC ever as more than 440,000 learners convene for English language learning*. [Accessed: 20-04-2020]. URL: https://www.futurelearn.com/info/press-releases/futurelearn-delivers-the-largest-mooc-ever-as-nearly-400000-learners-convene-for-english-language-learning.

[7] Pat Alvarado. *Learning Paths: Accelerating the Speed to Proficiency*. [Accessed: 20-04-2020]. URL: https://www.linkedin.com/pulse/learning-paths-accelerating-speed-proficiency-pat-alvarado/.

[8] Ry Rivard. *Measuring the MOOC Dropout Rate*. [Accessed: 20-04-2020]. URL: https://www.insidehighered.com/news/2013/03/08/researchers-explore-who-taking-moocs-and-why-so-many-drop-out.

[9] Shobita Thomas. *New on Coursera: start-to-finish learning paths for starting a new career*. [Accessed: 20-04-2020]. URL: https://blog.coursera.org/new-coursera-start-finish-learning-paths-starting-new-career/.

[10] Dan Davis et al. "Gauging MOOC Learners' Adherence to the Designed Learning Path". In: *EDM '16: 9th International Conference on Educational Data Mining*. 2016, pp. 54–61.

[11] Kurt Vanlehn et al. "The Andes Physics Tutoring System: Five Years of Evaluations." In: vol. 15. Jan. 2005, pp. 678–685.

[12] David H. Shanabrook et al. "Identifying High-Level Student Behavior Using Sequence-based Motif Discovery". In: *Educational Data Mining 2010, The Third International Conference on Educational Data Mining, Pittsburgh, PA, USA, June 11-13, 2010. Proceedings*. www.educationaldatamining.org, 2010, pp. 191–200.

[13] H. Chuang and C. Shen. "A Study on the Applications of Learning Paths Concepts to the Teaching in Elementary School". In: *2008 Eighth International Conference on Intelligent Systems Design and Applications*. Vol. 2. 2008, pp. 543–548.

[14] Bill Chiu, Eamonn Keogh, and Stefano Lonardi. "Probabilistic Discovery of Time Series Motifs". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Mar. 2003). DOI: 10.1145/956750.956808.

[15] Vequrity Inc. *PRIVACY POLICY OF RANGEFORCE*. [Accessed: 20-04-2020]. URL: https://hub.rangeforce.com/privacy-policy.

[16] General Data Protection Regulation (GDPR) Compliance Guidelines. *What is considered personal data under the EU GDPR?* [Accessed: 20-04-2020]. URL: https://gdpr.eu/eu-gdpr-personal-data/.

[17] National Institute of Standards and Technology. *National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework*. [Accessed: 20-04-2020]. URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-181.pdf.

Appendices

Appendix 1 - RangeForce modules as of February 5, 2020

CSCUK17 - Something is Calling Me	WASE Challenge - Delta LVL3
CSCUK17 - Something is Watching	WASE Challenge - Delta LVL4
CSCUK17 Face2Face - Something is Wrong (Leader)	WASE Challenge - Gamma LVL1
CSCUK17 Face2Face - Something is Wrong (Member)	WASE Challenge - Gamma LVL2
CSCUK18 Masterclass - Cryptofactor (Member)	WASE Challenge - Gamma LVL3
CSCUK18 Masterclass - Cryptofactor Infra	WASE Challenge - JWT 1
CSCUK19 - Qualifier 2	WASE Challenge - JWT 2
CSCUK19 Face2Face - API Bleed (Member)	WASE Challenge - JWT 3
CSCUK19 Face2Face - API Bleed Infra	WASE Challenge - Kappa
DevOps - Docker Dockerfile	WASE Challenge - Phone
DevOps - Docker Introduction	WASE Challenge - Spider
DevOps - Docker Networking	WASE Game
DevOps - File Management	WASE Learning - API Security: Exposed Tokens
DevOps - HTTPS Security	WASE Learning - Blind Command Injection: Find & Exploit (NodeJS)
DevOps - HTTPS Security: Apache	WASE Learning - Blind Command Injection: Fix (NodeJS)
DevOps - HTTPS Security: Introduction	WASE Learning - Blind NoSQL Injection: Find & Exploit (Meteor)

Continues...

Table 4 – *Continues...*

DevOps - HTTPS Security: Nginx	WASE Learning - Blind SQL Injection: Find & Exploit (PHP)
DevOps - Introductory Lab	WASE Learning - Command Injection [DEPRECATED]
DevOps - Software Management	WASE Learning - Command Injection: Find & Exploit (PHP)
DevOps - User Management	WASE Learning - Command Injection: Fix (PHP)
DevSecOps - IDS/IPS Lab 1	WASE Learning - Cookie Security: HttpOnly [DEPRECATED]
Hackerloo	WASE Learning - Cookie Security: HttpOnly: Find & Exploit (NodeJS)
M365 Security Lab - Heroes vs Villains	WASE Learning - Cookie Security: HttpOnly: Find & Exploit (PHP)
M365 Security Lab - Heroes vs Villains 1902	WASE Learning - Cookie Security: HttpOnly: Fix (NodeJS)
MS - Microsoft Security Defender ATP	WASE Learning - Cookie Security: HttpOnly: Fix (PHP)
MS Enterprise Security Hands-on Lab	WASE Learning - Cookie Security: Secure [DEPRECATED]
MS Enterprise Security Hands-on Lab - Walkthrough	WASE Learning - Cookie Security: Secure: Find & Exploit (NodeJS)
MS Ready Hacker Lab	WASE Learning - Cookie Security: Secure: Find & Exploit (PHP)
Platform Test Lab	WASE Learning - Cookie Security: Secure: Fix (NodeJS)
Security Tools - Brute-force Defense	WASE Learning - Cookie Security: Secure: Fix (PHP)
Security Tools - Malware Analysis	WASE Learning - CSRF [DECOMISSIONED]
Security Tools - Nikto	WASE Learning - CSRF Defend [DECOMISSIONED]
Security Tools - Nmap: Basics	WASE Learning - DOM-based XSS
Security Tools - Nmap: SSH Enumeration	WASE Learning - DOM-based XSS: Find & Exploit
Security Tools - NoSQLMap	WASE Learning - Insecure Deserialization

Continues...

Table 4 – *Continues...*

Security Tools - Password Cracking	WASE Learning - Insecure Direct Object References [DEPRECATED]
Security Tools - Password Cracking 2	WASE Learning - Insecure Direct Object References 2: Exploit (PHP)
Security Tools - PCAP Forensics	WASE Learning - Insecure Direct Object References 2: Fix (PHP)
Siege of Utopia	WASE Learning - Insecure Direct Object References: Find & Exploit (NodeJS)
SOC - Botnet Takedown	WASE Learning - Insecure Direct Object References: Find & Exploit (PHP)
SOC - Docker RunC Container Escape	WASE Learning - Insecure Direct Object References: Fix (NodeJS)
SOC - IDS/IPS: Suricata Basics	WASE Learning - Insecure Direct Object References: Fix (PHP)
SOC - IDS/IPS: Suricata IDS Rules	WASE Learning - JSON Web Token Security
SOC - IDS/IPS: Suricata Rule Management	WASE Learning - NoSQL Injection 1: Exploit
SOC - Privilege Escalation: Docker Group	WASE Learning - NoSQL Injection 1: Find
SOC - Privilege Escalation: Kernel Exploit (Dirty Cow)	WASE Learning - NoSQL Injection 1: Fix
SOC - Privilege Escalation: LXD Group	WASE Learning - NoSQL Injection 2: Exploit
SOC - Privilege Escalation: Misconfigured PATH	WASE Learning - NoSQL Injection 2: Fix
SOC - Privilege Escalation: SUID Bit 1	WASE Learning - Path Traversal [DEPRECATED]
SOC - Privilege Escalation: SUID Bit 2	WASE Learning - Path Traversal: Find & Exploit (NodeJS)
SOC - Privilege Escalation: SUID Bit 3	WASE Learning - Path Traversal: Find & Exploit (PHP)
SOC - Privilege Escalation: Wildcard Injection	WASE Learning - Path Traversal: Fix (NodeJS)
SOC - Regular Expressions: Basic	WASE Learning - Path Traversal: Fix (PHP)

Continues...

Table 4 – *Continues...*

SOC - Regular Expressions: Intermediate	WASE Learning - SQL Injection [DEPRECATED]
SOC - Security Introduction	WASE Learning - SQL Injection: Authentication Bypass
SOC - System Compromised	WASE Learning - SQL Injection: Prelude
SOC Challenge - Backdoor 1	WASE Learning - SQL Injection: Union Select
SOC Challenge - Backdoor 2	WASE Learning - Unrestricted File Upload [DEPRECATED]
SOC Challenge - Backdoor 3	WASE Learning - Unrestricted File Upload: Find & Exploit (NodeJS)
SOC Challenge - Beta	WASE Learning - Unrestricted File Upload: Find & Exploit (PHP)
SOC Challenge - Fl4gPrint3r	WASE Learning - Unrestricted File Upload: Fix (NodeJS)
SOC Challenge - Privilege Escalation: SUID Bit	WASE Learning - Unrestricted File Upload: Fix (PHP)
SOC Challenge - Uncontained	WASE Learning - XML External Entities
SOC Challenge - Webmin	WASE Learning - XML External Entities 2
Trapped	WASE Learning - XML External Entities: Find & Exploit
Vulnerable Web Applications	WASE Learning - XSS Filter Evasion: Find & Exploit
WASE Assessment	WASE Learning - XSS Filter Evasion: Fix
WASE Challenge - Alpha LVL1	WASE Learning - XSS: Reflected
WASE Challenge - Alpha LVL2	WASE Learning - XSS: Stored
WASE Challenge - Alpha LVL3	WASE Learning - XSS: Stored-based Phishing
WASE Challenge - Delta LVL1	Windows Defender ATP Workshop
WASE Challenge - Delta LVL2	