Preliminary Project Report

Research and prototyping of an email phishing emulation system

Members, Group 23.

Martin Bang,	s315602.
Frithjof Brate,	s336406.
Kristian Marison Haugerud,	s331048.
Mohammad Tayyab Khalid,	s319229.
Jørgen Sandnes,	s331423.

Website: Bachelor-G23

Table of Contents

Presentation

Project Description

Research and Prototype of an email phishing emulation system that can be utilized to detect human weaknesses in an organization and estimate the cost of the breach. The phishing emails template should be based on the research of effective techniques used in today's industry. Finally, a prototype will be created to analyze and present the data.

Customer

CYBR is a corporation whose business is within the field of Cyber Security.
Their goals are, among others, to gamify security training to increase awareness of potential threats within the industry. They also provide security tools that scan for weaknesses within their customers' websites and network configurations and servers. The collected data of training and threats can then be reported back to the customers, who then can decide on necessary precautions.

Customer Contacts

CYBR,	Hovfaret 13, 0275 Oslo,	https://cybr.no
Joe Jones,	Director of Operations,	joe@cybr.no
Jaan Kitsuk,	Project Supervisor,	jaan@cybr.no
Prahkar Srivastav,	Technical Supervisor,	prakhar@cybr.no

Internal supervisor

Dr. Lothar Fritsch,	Professor of Applied Cybersecurity. Department of Computer Science, Oslo Metropolitan University.	lothar.fritsch@oslomet.no

Summary

CYBR's goal is to reinvent cybersecurity training by providing a gamified training platform. Therefore, the task is to do research and make a prototype of an email phishing emulation system. This system shall be used to check how vulnerable their customers are for phishing attacks and to estimate the potential cost of a breach.

We will research the most prominent phishing attack methods, email and landing page templates. This research will be done by contacting major actors in the Norwegian cybersecurity branch and gather literature from Google Scholar and Oria. The results will be cross-referenced, sorted, and analyzed. All information gathered will be used to make efficient phishing email and landing page templates.

The backend solution will utilize the REST API of the phishing framework GoPhish. A Python API client will be used to extract and analyze data to be stored in a database. Here we can sort phishing targets into groups, import email and landing page templates, and launch phishing campaigns. Email opened, links clicked and data harvested will be displayed in a minimalistic GUI.

Our team will adapt an agile workflow with elements from SCRUM and Kanban adjusted to our project size. A progress plan has been made to keep track of major milestones. We have based us on a remote work environment because of COVID-19 with tools such as Slack, Trello and Google Docs.

Current situation

CYBR is very heavily involved in commercialisation of its newly released product. Due to this, the resources they can use for research are limited. Hence, they want to use us as a research team to help them improve their Phishing Simulation product, which is not released on the market yet.

They want to help their customers remove the need for consultants to help them with cybersecurity and automate the cybersecurity optimisation process. Currently, the main focus of the CYBR Platform is to create a gamified cybersecurity training platform, where each employee has a tailored training program. The training program is generated based on each individual employee's assessments, simulated phishing emails fail rates, and other KPIs.

The company goal is to release a market leading Phishing Simulation product during the first half of 2021. This product should check how good the employees are at detecting phishing emails and it should calculate the financial risk per employee based on the position, industry, and Phishing KPIs . Moreover, that is what we are going to make.

Specifications for the research

Prerequisites

- The researched data is a cornerstone of the project.
- Research data should be used to establish how to implement and how the system behaves.
- Phishing framework needs to be open-source.
- The research should evaluate the financial risk of the breach via email per each individual employee.

Core tasks

- Create a list of studies that could be used in a corporate setting.
- An excel document that lists articles over phishing attacks we have researched. Including how costly it was, records of the breach, affected companies, type of attacks, and a description.
- What email phishing attacks were most successful and why.
- Understanding and evaluating modus operandi for phishing attacks.
- Identify an attack mechanism that can be used as a pipeline to make the attacks more likely to succeed if the first try fails.

Specifications for the technical implementation

Requirements

- Phishing servers will be whitelisted in firewalls bypassing any security measures enforced by companies.
- Open-sources phishing framework as the backbone of our project.
- Frontend should be a minimalistic GUI for data presentation.
- It should be easy to import email and or other types of information
- A showcase of statistics for the different campaigns that were used.
- The possibility to create custom user groups.
    - Ability to create premade groups as well based on position and departments.
    - Groups for each company.
- A database with premade email templates.
    - Use the most efficient templates based on the research.
- A database with premade landing pages based on the research data.
- Store metric data over different campaigns separately for each company/department.
    - Emails sent.
    - Emails accepted.
    - Emails delivered.
    - Emails opened.
    - Links clicked.
    - Emails reported.
    - Data gathered.
- An option that allows for scheduling before a campaign is launched.
- Final product should be containerized.

Additional features (if time)

- Long term engagement with high priority targets.
- A grading system (based on how many clicked the link, which can be helpful later when they are going to calculate the costs. e.g., Risk level.)
- Email import: import from other management tools, such as azure active directory

Research Solution

Information gathering

To create a solid foundation for the project to be built upon, we need to gather appropriate information. We will contact institutions in Norway and try to get data on widely used methods in today's phishing industry. The articles and reports will be sorted into different attacks, the attack scope, and the cost of the breach. The results will be cross-referenced and used to generate templates for our phishing emulation system. We can then estimate the cost for the different attack methods.

Sources of information

Relevant actors

The Norwegian Center for Information Security	https://norsis.no/english/
The Norwegian Business and Industry Security Council	https://www.nsr-org.no/english
Norwegian National Security Authority	https://nsm.no/home/
Nettvett - Etiquette in technology	https://nettvett.no/

Other:

Anti-Phishing Working Group	https://apwg.org/
OsloMet University Library - Oria	https://www.oslomet.no/en/ul
Google Scholar	https://scholar.google.com/

Template creation

The cross-referenced results will be sorted, and the most prominent attack methods will be analyzed. These will be broken up and sorted into categories. The common denominator will be extracted and used to create templates and landing pages within each category. Then they will be implemented into the technical solution.

Estimate cost of the breach

Data from analyzed attacks will be stored and used to estimate a cost for other attack methods. By gathering information on those methods, credentials harvested, high-risk targets, and the cost of the breach within the genres, we can estimate each of the research templates' cost.

Technical solution

Overview

In coming up with a technical solution for our project, various frameworks and tools were considered. The idea is to create a back-end service that can interact with a phishing simulator's API and display the data from our backend to a custom frontend. In the frontend, the information will be displayed in an organized manner. It will contain the statistics from the launched campaigns.

As provided in this attachment is an overview of the project's architectural design:
Attachment 2: High-Level Design

CYBR wants us to put most of our attention into the backend, as that is where the most important aspects of the project will be covered. Regarding the frontend, CYBR has provided an inspirational design:
Attachment 3: Front End Inspiration - Dashboard
Attachment 4: Front End Inspiration - New Campaign

Platform

The whole project will be stored and operated on a Google Cloud developer platform provided by CYBR. On the platform, CYBR has set up a ubuntu virtual environment (Canonical, Ubuntu, 20.04 LTS) for us to utilize, where we can develop the necessary docker containers for the clients and the servers. The final project will be shipped separately as one frontend and backend container.

Languages

Python is used for our backend service and javascript for our frontend. Both are used at CYBR, and we thought that sticking with the same choices would be more beneficial for them. The backend will utilize Python API clients.

Frameworks

We went with the framework Django for our backend, because it is an efficient framework for building a REST API. Over to the frontend, our decision landed on React, because of how well it corporates with Django. With React we can easily call API-endpoints from our backend and display it using minimal amounts of code. Attached is a diagram with the frameworks in use:
Attachment 5: UML - Project Setup

The last and most important framework is the phishing tool. It is the backbone of our project. We ended up with Gophish, a framework that could perform many tasks needed for our project and have a solid structured RESTful API that we can utilize.
Some charts further explaining the flow of the GUI:
Attachment 6: Flow chart - Watch customer statistics
Attachment 7: Flow chart - Launch campaign

Possible outcomes

This project's primary goal is to gather information about the most effective phishing attacks and deliver a minimal viable product of a phishing emulation system to CYBR. The research will improve phishing email templates and landing pages.

If the product fits the demands CYBR has for technical specifications, automation, and efficiency. Parts of the product can be implemented in their existing cybersecurity tool. If not, CYBR will have gathered valuable research.

CYBR wants to have a product that is easy to use. Importing groups, selecting templates, and launching campaigns should be as automated as possible. We potentially do not have enough time to implement and refine this during the project within the timeframe.

While going through the tools and frameworks used in this project, it is important to note that these might change during development and that challenges might occur, causing us to change paths.

Work Schedule

Phase 1: Planning

Project Outline

A project outline is a document that we must deliver before we can continue with our work. In this report, we have to define our project so the Oslo Metropolitan University can approve before moving on with our work. It should also contain the contact information for the client and us. It is an overview of the initial project.

Website

We made a pretty basic website where we can post who we are and the various documents we have to hand in. Link to website: Bachelor-G23

Preliminary project report

The preliminary report is an overview of the work that already is done and what we are going to do moving forward. It shall work as a description between us, CYBR and the university.

Work and project schedule

These documents shall be handed in together with the Preliminary Report. The work schedule should contain an overview of our work. The project schedule shows when we need to do different activities. For example, we work with the "Preliminary Report" in weeks 2 and 3.

Phase 2: Research

Research

The research part is in several stages. Firstly, we have to get a solid understanding of what phishing is and how critical phishing attacks can be. Afterwards, our research needs to be more specific. For example, which technologies are used to make phishing attacks.

Requirements

To ensure that the project provider and our team are on the same page.

Phase 3: Implementation

Development

Here we use our requirements and research to develop the product our client wants. Before starting the development, we must know which technologies we are going to use. This will be defined in phase 2.

Create templates

We need to make the email templates that are going to be used in our simulated phishing attacks. These templates should be made by looking at the research we have done. For example, what type of phishing emails are most effective in different situations.

Phase 4: Testing

Testing

In this phase the product gets tested. This is done to ensure that the finished product fulfills the requirements.

Phase 5: Documentation

Final report

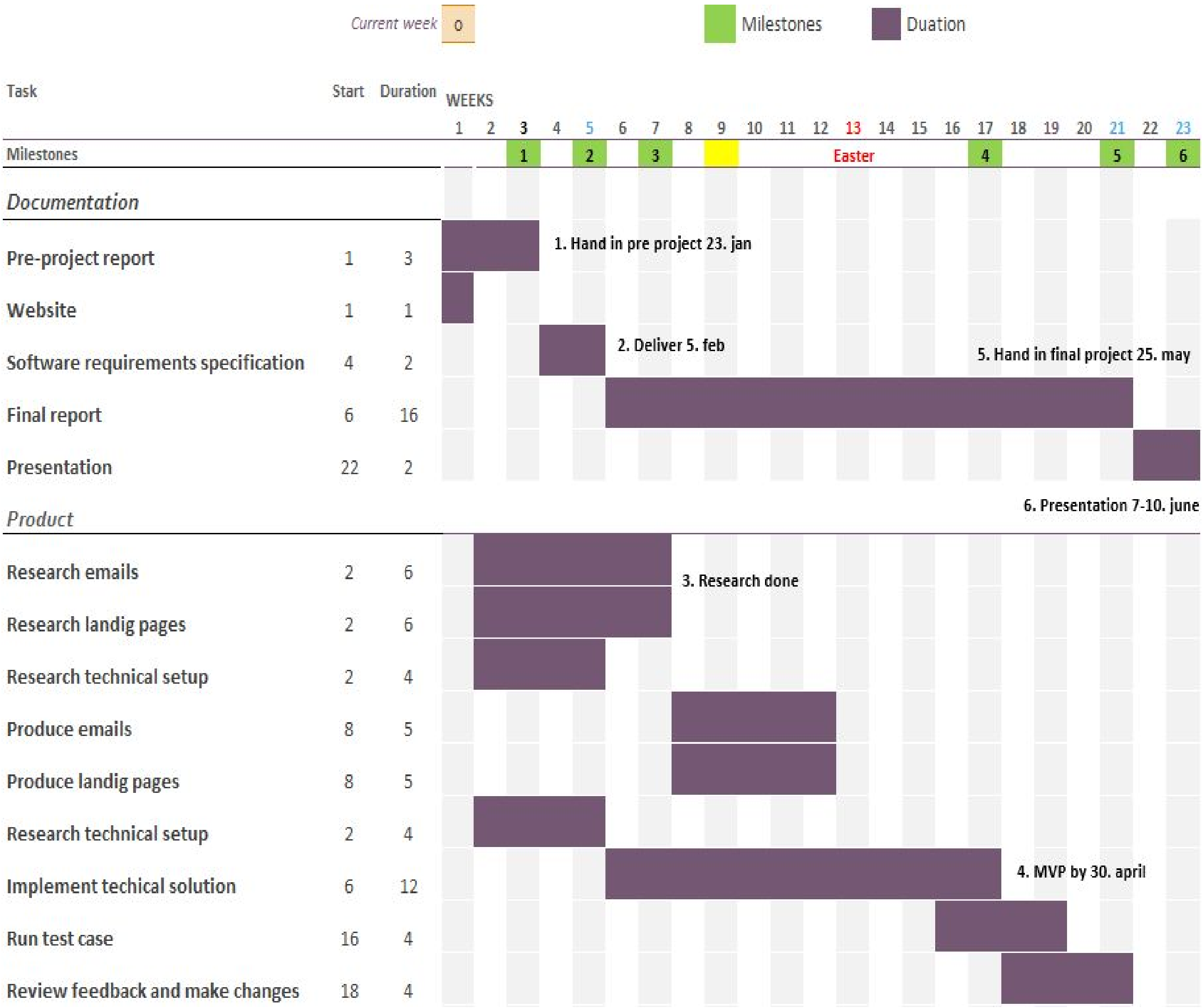
The final report documents the whole working process. It must contain how and when we have worked also our final results where it must tell our thoughts about the decisions we made.

Presentation

A presentation to show off our work and results.

Project Schedule

The Gant scheme serves as an initial high-level project overview. Here we have marked all milestones and can instantly see what work is most important. We will also use it to keep track of how far we have progressed with every task. If necessary, we can adjust if we finish a task early or allocate more time if needed. The work on the documentation for the final report will be worked on through the whole project duration. Then we can use the last week to improve and review our report. Any extra time we have leftover can be used on coding and implementation.



Task	Start	Duration
Documentation		
Pre-project report	1	3
Website	1	1
Software requirements specification	4	2
Final report	6	16
Presentation	22	2
Product		
Research emails	2	6
Research landig pages	2	6
Research technical setup	2	4
Produce emails	8	5
Produce landig pages	8	5
Research technical setup	2	4
Implement techical solution	6	12
Run test case	16	4
Review feedback and make changes	18	4

Work methods

We will implement an agile workflow and use elements from SCRUM and Kanban that fits our project size. Our team members have had success with an agile workflow in earlier projects at OsloMet.

SCRUM will be used as a guide and at the start of the project. We will have daily stand-ups during the weekdays. Code demos and a review of work done will be held as we see fit. Many meetings can be interruptive to the personal workflow. Nonetheless/however, we believe itis important that everyone is on the same page which ensures progress

Sprints will be planned after the most important requirement from the customer, and we will keep a sprint backlog. A sprint review with CYBR will be held after each completed sprint and sprint retrospectives will be used to improve internal working methods.

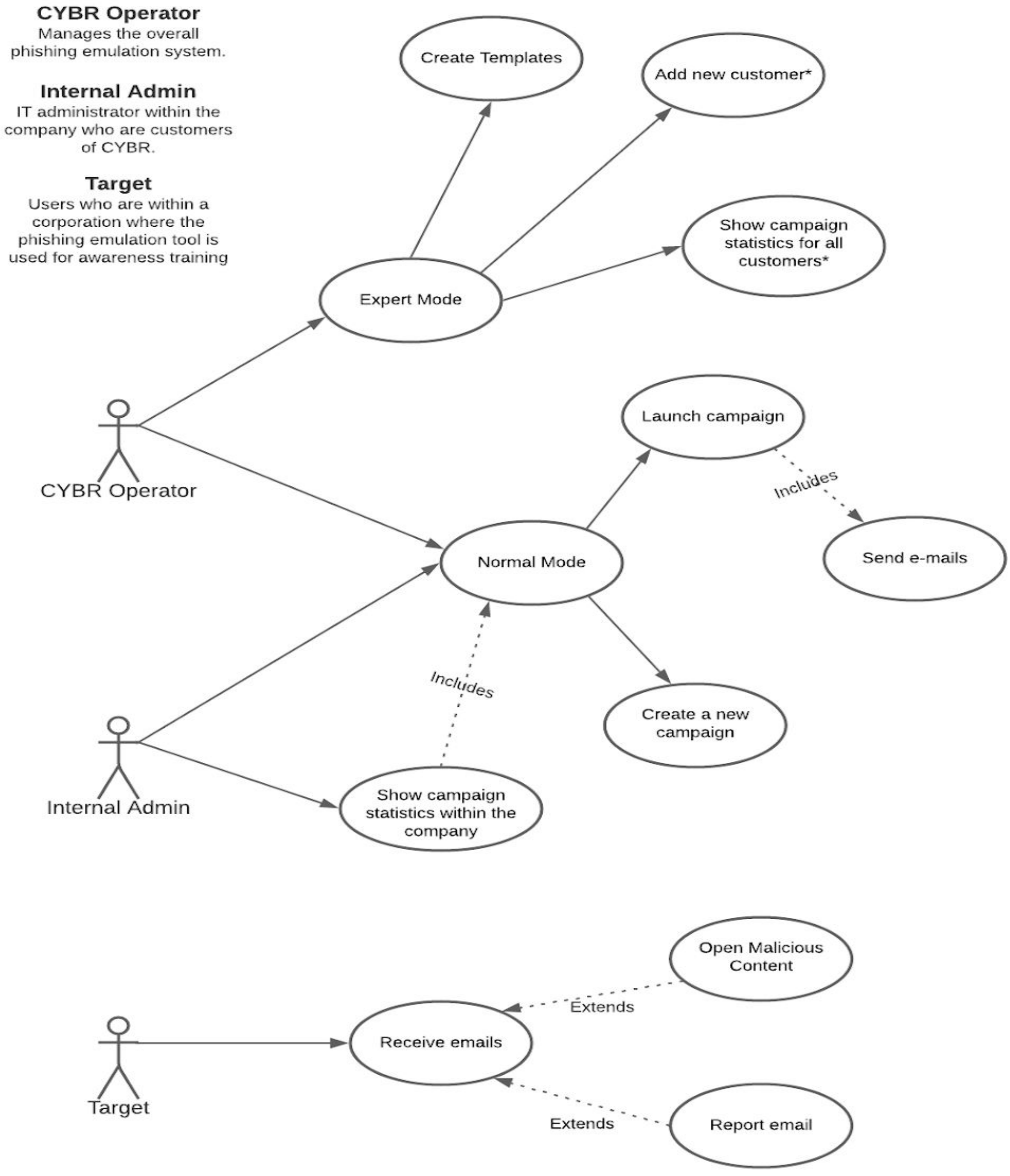
Daily tasks are divided in the team, and the same goes for further responsibility. A scrum master and product owner has been assigned. The plan is the stick with these roles for the whole project so we can be efficient and learn their respective roles

Tasks will be assigned on a kanban board. Here we have a clear overview of who is doing what, the due date on tasks, and the priority the tasks should be completed. Every task will be deconstructed into smaller problems and put on a Trello board card. Cards are used to track progress and problems. Every team member then has a clear overview of the work done and can easily get up to date on the problem. Easy to assist if any problems should arise.

We base our workflow on remote work due to the COVID-19 coronavirus regulations. Slack is used daily to resolve any problems fast and host all meetings. We also have a slack channel with our customer CYBR for fast communication.

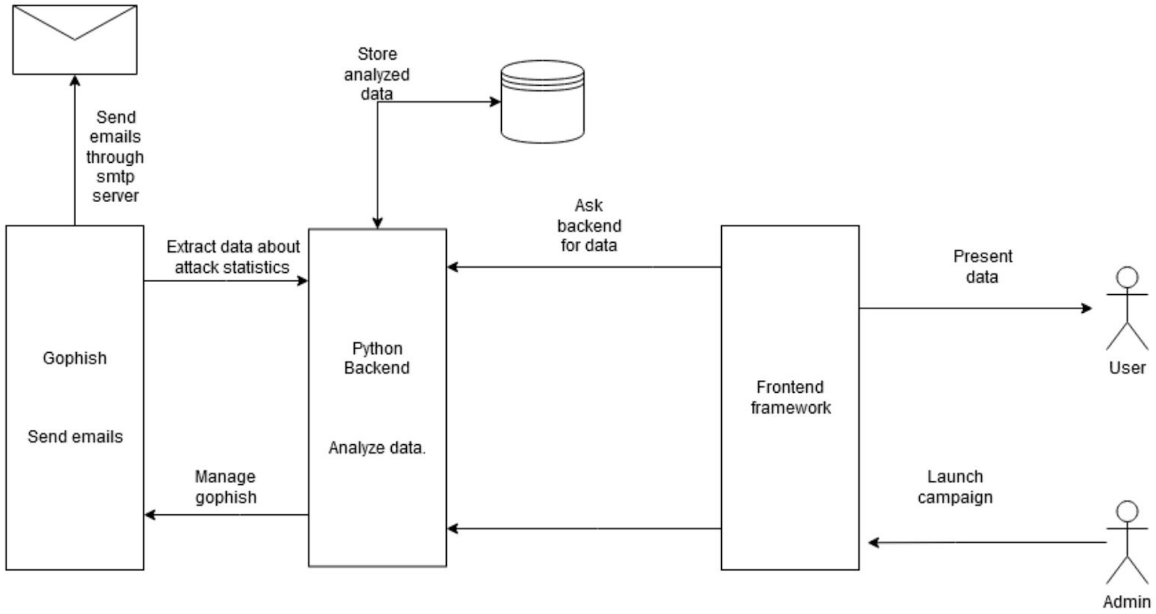
Attachments

Attachment 1: Use Case - Diagram


CYBR Operator
Manages the overall phishing emulation system.
Internal Admin
IT administrator within the company who are customers of CYBR.
Target
Users who are within a corporation where the phishing emulation tool is used for awareness training
Create Templates
Add new customer*
Show campaign statistics for all customers*
Expert Mode
Launch campaign
Includes
Send e-mails
CYBR Operator
Normal Mode
Includes
Create a new campaign
Internal Admin
Show campaign statistics within the company
Open Malicious Content
Extends
Receive emails
Target
Extends
Report email

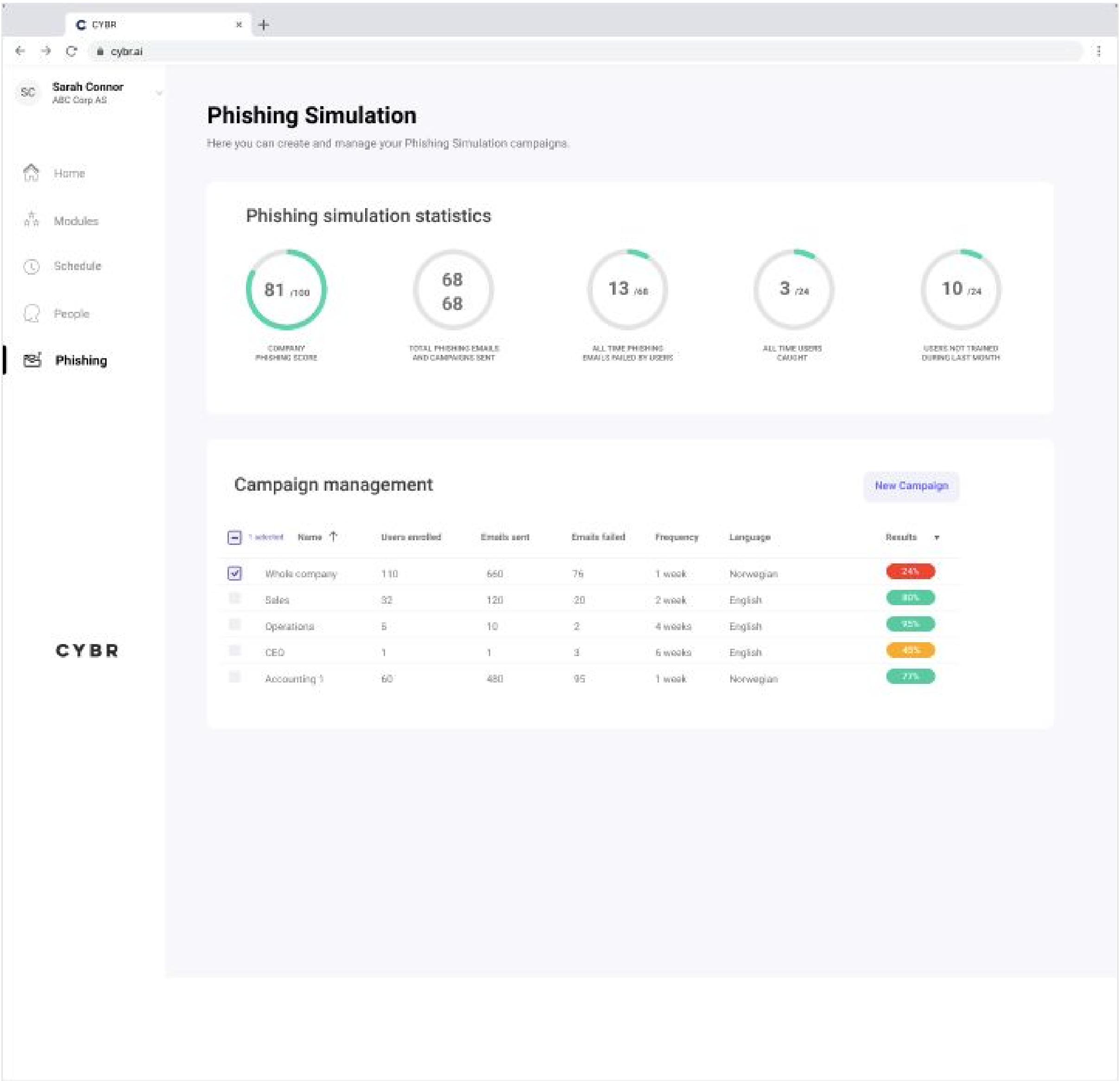
* Customers are referred to as customers of CYBR.

Attachment 2: High-Level Design



Attachment 3: Front End Inspiration - Dashboard

Inspiration made by CYBR:



Phishing Simulation

Here you can create and manage your Phishing Simulation campaigns.

Phishing simulation statistics

81 /100 COMPANY PHISHING SCORE

68 68 TOTAL PHISHING EMAILS AND CAMPAIGNS SENT

13 /68 ALL TIME PHISHING EMAILS FAILED BY USERS

3 /24 ALL TIME USERS CAUGHT

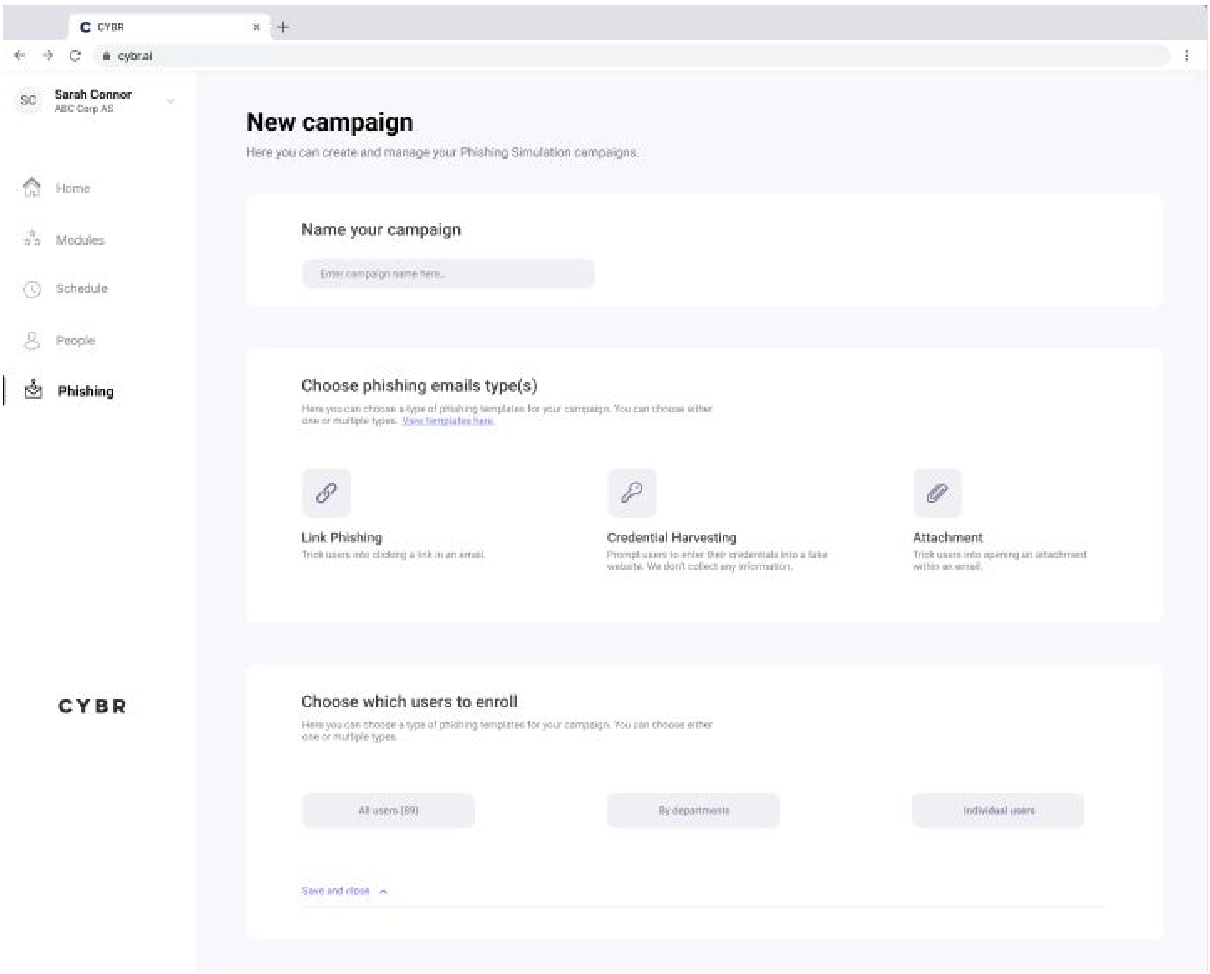
10 /24 USERS NOT TRAINED DURING LAST MONTH

Campaign management

⊟ 1 selected	Name ↑	Users enrolled	Emails sent	Emails failed	Frequency	Language	Results ▾
☑	Whole company	110	660	76	1 week	Norwegian	24%
□	Sales	32	120	20	2 week	English	80%
□	Operations	5	10	2	4 weeks	English	95%
□	CEO	1	1	3	6 weeks	English	45%
□	Accounting 1	60	480	95	1 week	Norwegian	77%

Attachment 4: Front End Inspiration - New Campaign

Inspiration made by CYBR:



New campaign

Here you can create and manage your Phishing Simulation campaigns.

Name your campaign

Enter campaign name here...

Choose phishing emails type(s)

Here you can choose a type of phishing templates for your campaign. You can choose either one or multiple types. View templates here.

Link Phishing
Trick users into clicking a link in an email.

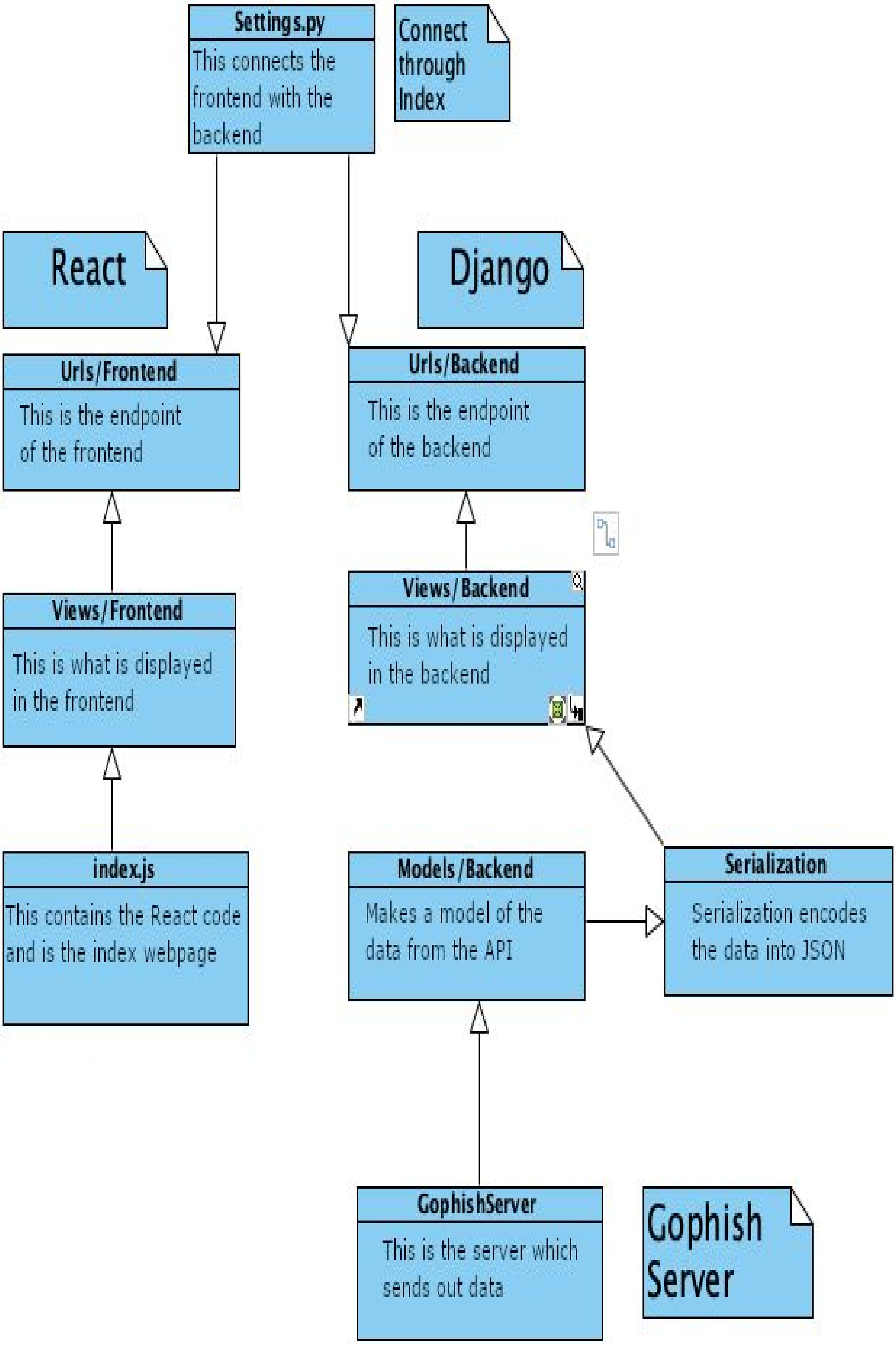
Credential Harvesting
Prompt users to enter their credentials into a fake website. We don't collect any information.

Attachment
Trick users into opening an attachment within an email.

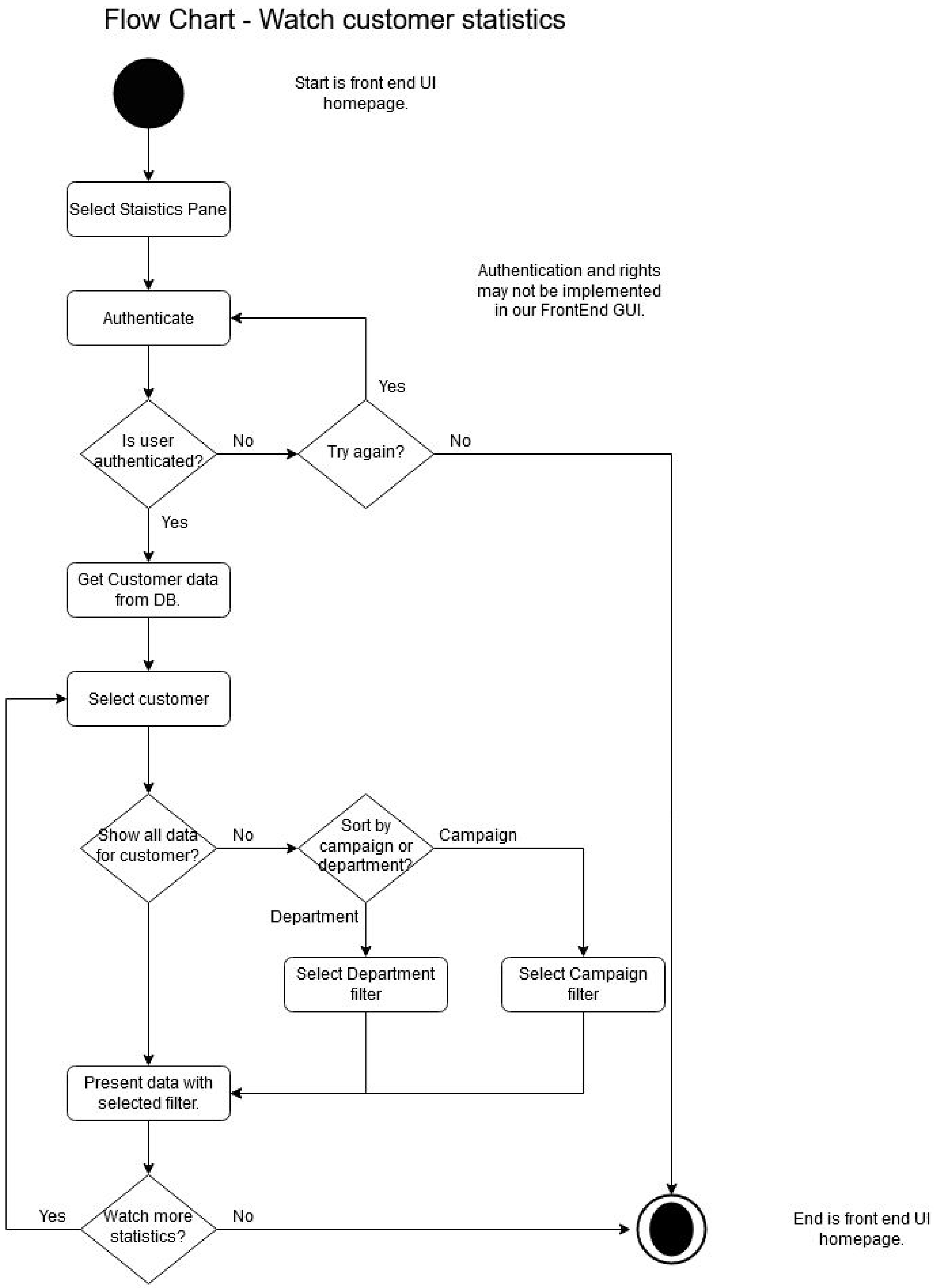
Choose which users to enroll

Here you can choose a type of phishing templates for your campaign. You can choose either one or multiple types.

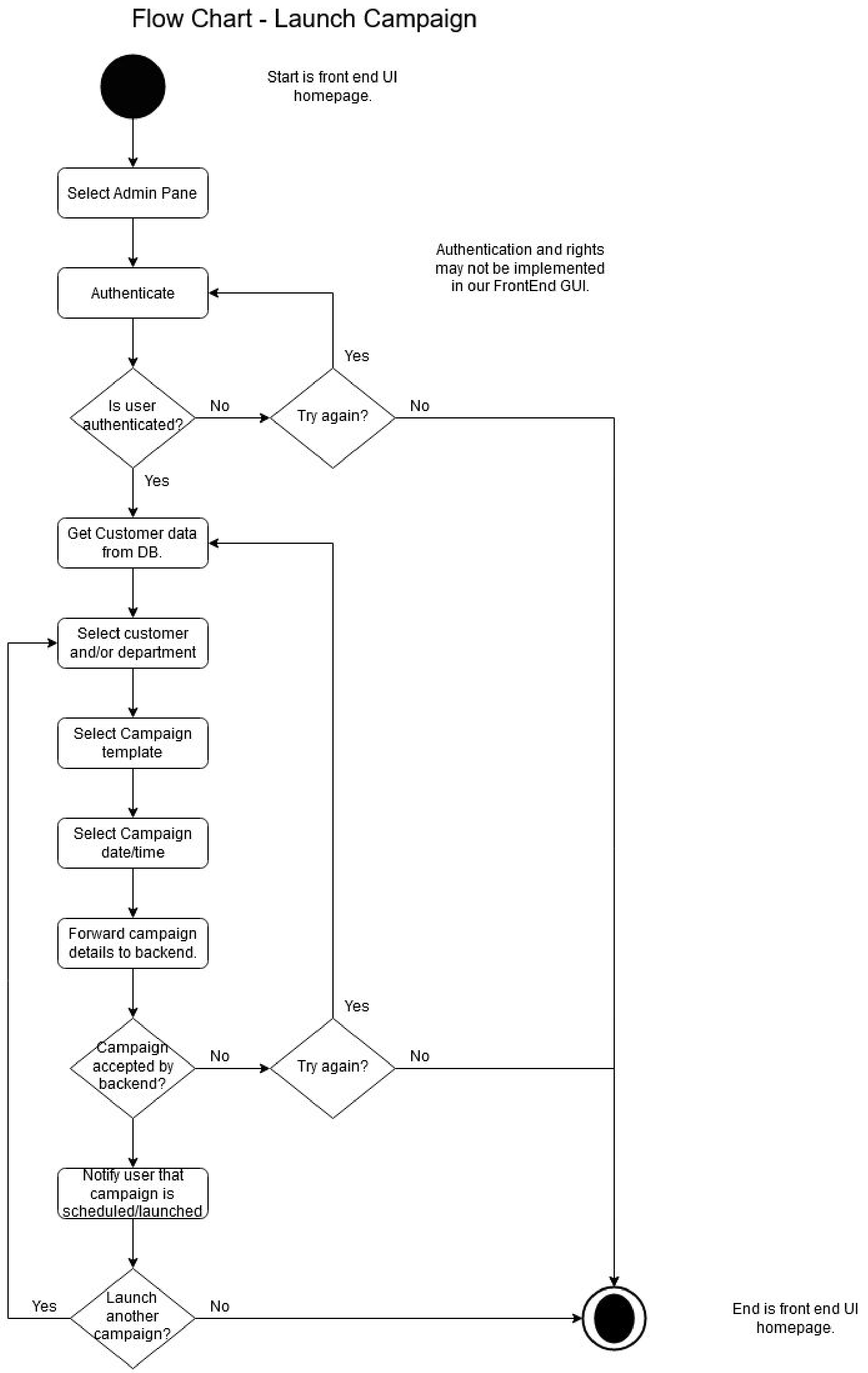
Attachment 5: UML - Project Setup


Settings.py
This connects the frontend with the backend
Connect through Index
React
Django
Urls/Frontend
This is the endpoint of the frontend
Urls/Backend
This is the endpoint of the backend
Views/Frontend
This is what is displayed in the frontend
Views/Backend
This is what is displayed in the backend
index.js
This contains the React code and is the index webpage
Models/Backend
Makes a model of the data from the API
Serialization
Serialization encodes the data into JSON
GophishServer
This is the server which sends out data
Gophish Server

Attachment 6: Flow chart - Watch customer statistics


Flow Chart - Watch customer statistics
Start is front end UI homepage.
Select Staistics Pane
Authenticate
Authentication and rights may not be implemented in our FrontEnd GUI.
Yes
Is user authenticated?
No
Try again?
No
Yes
Get Customer data from DB.
Select customer
Show all data for customer?
No
Sort by campaign or department?
Campaign
Department
Select Department filter
Select Campaign filter
Present data with selected filter.
Yes
Watch more statistics?
No
End is front end UI homepage.

Attachment 7: Flow chart - Launch Campaign


Flow Chart - Launch Campaign
Start is front end UI homepage.
Select Admin Pane
Authenticate
Authentication and rights may not be implemented in our FrontEnd GUI.
Is user authenticated?
No
Try again?
Yes
No
Yes
Get Customer data from DB.
Select customer and/or department
Select Campaign template
Select Campaign date/time
Forward campaign details to backend.
Campaign accepted by backend?
No
Try again?
Yes
No
Notify user that campaign is scheduled/launched
Yes
Launch another campaign?
No
End is front end UI homepage.