university of groningen

faculty of science and engineering

Collecting Data from exposed digital twins in offshore wind farms

From discovery to characterisation: roles, potential digital twins, and one-month change in exposed services near offshore wind farms

Bachelor's Project Computing Science

August 2025

Author: Vasco Reynolds Brandao
Student Number: S5072298
First supervisor: Dr Fadi Mohsen
Second supervisor: Dr Dilek Dustegor

Abstract

The convergence of IT and OT in offshore wind has heightened concerns where insecure or legacy protocols remain in use. Digital twins are increasingly applied across wind-farm assets and when underlying systems exhibit cybersecurity weaknesses, those risks can be reflected in digital twins as well. This study focuses on European offshore wind farms and on devices whose exposed services use selected insecure or legacy protocols (DNP3, FTP, Modbus, Siemens S7, Telnet). Its aims : to identify devices plausibly related to offshore wind farms and assign them a role within the wind-farm context (Turbine, Substation, Control Center, Other, or Not Part of Offshore Wind Farms(OWF)), to flag potential digital twins, to observe short-term changes in selected service-level fields and describe Common Vulnerabilites and Exposures(CVEs) associated with the devices' services. The potential digital-twin identification and change-over-time observation are applied only to devices first attributed an OWF role. To support reuse and extension, the work provides a modular python CLI and structured data/criteria files designed so other researchers can replicate or adapt the investigation. The study contributes a transparent classification system for roles and potential twins, together with a compact observations of short-term changes in service information and CVEs.

Contents

List of Figures

List of Tables

1 Introduction & Motivation

With 74 active offshore wind farms, 12 in construction, 14 approved and 491 planned, Europe has become one of the global leaders in offshore wind [5]. According to Statista, Over the past 13 years, Europe's cumulative offshore wind capacity installations has gone from 5 GW in 2012 to 37 GW in 2024[6]. European investments in this particular energy sector are predicted to exceed USD 40 B $ between 2024 and 2025 [7] alone. In the pursuit for lower operational costs and efficiency, the convergence of Information Technology (IT) and Operational Technology (OT) in critical infrastructure, including offshore wind farms, has accelerated. In return this raises serious concerns with regards to cyber security given that OT was not integrated with IT when a lot of industrial systems were first developed. Hence lack of authentication and encryption of data wasn't a concern as these systems were safe by being isolated from IT[12].

In the case of offshore wind farms specifically, this convergence is necessary to enable remote operations and diagnostics for example, which given the remote environments of offshore wind farms is beneficial as it saves on operations and maintenance costs. Recent sector incidents include ENERCON's loss of remote access following the KA-SAT modem wiper event and ransomware events at Nordex and Deutsche Windtechnik, which disrupted remote operations[20].

Another benefit brought by the convergence of OT and IT in order to improve performance and maintenance is the use of digital twins. Digital twins are digital replicas of devices and/or entire systems. One of the main uses of digital twins in the context of offshore wind farms is for predictive maintenance, where digital twins are used to predict failure of equipment which can reduce operations and maintenance (O&M) costs as 38% of offshore operating costs are allocated to maintenance[9]. In digital twin applications the concerns regarding cyber security due to lack of authentication and encryption persists, as depend on data interfaces that may reuse the same operational networks and protocols. If those interfaces lack strong authentication or encryption, similar risks can surface in digital twin contexts.

The issue of cyber security concerns in offshore wind farm environments due to the convergence of OT and IT is important to address as it could compromise Europe's (and the rest of the world's too) energy production sector. This issue not only affects the organizations who operate offshore windfarms, but also all the people who are reliant on the energy they generate. In order to address this issue it is important to explore what devices in the public internet with these potential vulnerabilities are exposed.

In this study we intend to contribute to research of cyber security concerns related to offshore wind farms by asking the questions:

1. Can we discover plausible offshore wind devices using insecure or legacy protocols in the public Internet?

2. Can we classify offshore wind devices found in the public Internet into roles based on

service metadata?

3. Can we identify potential digital twins within offshore wind devices found in the public Internet based on service meta-data?

4. What types of service data changes can be found in offshore wind devices found in the public Internet?

5. Can we find associated CVEs in offshore wind devices found in the public Internet?

2 Literature Review / State of the Art

2.1 Background

Offshore wind farms(OWFs) are large, distributed infrastructures that combine industrial control systems (ICS) with information technology (IT) platforms to manage energy production, transmission, and monitoring. Their key components include turbines, substations, and control centers. All these components are supported by supervisory control and data acquisition (SCADA) systems that connect OT assets with IT-based management platforms.

Turbines produce electricity and contain controllers and sensors that regulate blade pitch, yaw, performance, and other systems. These functions are coordinated by programmable logic controllers (PLCs) and turbine-level SCADA. Turbines are generally supplied and partially managed by vendors such as Siemens Gamesa or Vestas, while day-to-day operation is managed by the wind farm operator [2, 3].

Substations collect the electricity generated turbines and transfer it to the onshore grid. Offshore substations (OSS) host transformers, switchgear, and protection relays, all of which are controlled by ICS devices that communicate through industrial protocols such as DNP3 or IEC 60870-5-104. While operators typically oversee substation systems, vendors may remain responsible for specific protection and monitoring equipment [8, 3].

Control centers act as the operation centers of OWFs. They aggregate data from turbines and substations through SCADA systems, allowing operators to monitor performance and respond to alarms. Increasingly, these centers also integrate with corporate IT systems to support analytics, visualization, and predictive tools, often supplied by vendors as part of monitoring service contracts [17].

The main stakeholders in OWFs are operators, who are responsible for overall performance and profitability, and vendors, who supply and often retain management responsibilities for turbines, SCADA platforms, and some monitoring equipment.

Overall, the convergence of IT and OT in OWFs occurs where SCADA and ICS networks interface with IT systems for remote access, monitoring, and analytics. While this has many benefits, it also creates pathways where vulnerabilities in IT environments can compromise OT systems.

2.2 Security Incidents involving Offshore Wind Farms

Europe's offshore wind sector has become a critical pillar of the continent's energy transition. In 2023, offshore wind accounted for around 4% of the Europe's electricity consumption, and policy targets foresee offshore wind delivering at least 30% of Europe's power demand by 2050, equivalent to 400–450 GW of installed capacity [1]. However, recent years have shown that the sector is not immune to cyberattacks, underlining the importance of robust cybersecurity practices.

In Germany, Nordex, Deutsche Windtechnik, and Enercon, were affected by cyber incidents that shared a common consequence: the disruption of remote monitoring and control systems [4]. Nordex suffered a ransomware attack in March 2022, that forced shutdown of its IT systems, preventing operators from remotely supervising turbines. In April 2022, Deutsche Windtechnik was hit by a similar attack, causing its remote monitoring platform to go offline. In the same year, Enercon lost visibility over 5,800 turbines when the Viasat KA-SAT satellite network was disrupted. In all cases, turbine generation continued, but the loss of remote supervision increased risks for maintenance and safety. If attackers had escalated access, consequences could have included forced turbine shutdowns or even cascading failures across farms.

Other companies have faced different types of incidents. In 2021, Vestas experienced a ransomware attack that compromised corporate IT systems across multiple countries. Although turbines remained operational, sensitive internal data was stolen, raising concerns over intellectual property theft and reputational damage [4]. Although no OT systems were confirmed as impacted, the attack highlighted how corporate IT breaches could provide stepping stones into turbine control networks if not properly segmented.

In addition, Ørsted and other operators reported being subject to repeated phishing and social engineering attempts [4]. Successful compromises of this type could allow attackers to pivot into administrative environments, deploy malware, or bridge into OT systems, underscoring the persistent threat landscape.

2.3 The risks of using Insecure Protocols

In this section of the state of the art we discuss the risks of using legacy or insecure protocols in contexts such as offshore wind farms, where services running them could be vulnerable to cyberattacks. We chose to highlight five protocols due to their use in ICS and references in security related literature: DNP3, FTP, Modbus, Siemens S7, and Telnet. These protocols are used in communication in SCADA systems, substations, and turbine level controllers, and their weaknesses can directly extend to offshore wind farm operations.

DNP3 (Distributed Network Protocol) is widely used for communication between control centers, Remote Terminal Units (RTUs), and Intelligent Electronic Devices (IEDs). Its role in offshore wind farms includes transmitting telemetry data from substations and relays to control centers. However, the original specification lacks encryption and authentication, making it susceptible to spoofing or man-in-the-middle attacks if exposed to the internet. In the context of offshore wind farms, compromised DNP3 traffic could allow adversaries to send false telemetry or malicious commands, potentially leading to incorrect dispatch decisions or loss of visibility of farm output [8].

FTP (File Transfer Protocol) is used for transmitting logs, firmware, or configuration files between turbines, substations, and vendor maintenance systems. Its plain-text nature means that both authentication credentials and transferred files can be intercepted. In offshore wind settings, where FTP is sometimes used by vendors for maintenance data uploads or turbine firmware updates, attackers could replace files with malicious versions or harvest sensitive operational information, extending the attack surface through misconfigured or outdated systems [3].

Modbus, is one of the oldest ICS protocols and is still used to communicate with turbine controllers, sensors, and substation equipment. It was designed for simplicity and does not include security features such as authentication or integrity checks. In offshore wind farms, Modbus is used for polling sensor data like voltage, current, and rotor speeds. An attacker with access could manipulate information, causing turbines to operate outside of safe conditions or hiding away faults [13].

The Siemens S7 protocol is specific to Siemens programmable logic controllers (PLCs), which are often deployed in wind turbines and substations. S7 communications allow reading and writing directly to PLC memory and logic. Vulnerabilities in S7 have historically been exploited. In offshore wind contexts, insecure or exposed S7 channels could allow attackers to alter turbine control logic or disable protection mechanisms, creating risks of both physical damage and large-scale outages [8].

Telnet, although largely deprecated, persists in some vendor-supplied systems and legacy devices in wind farms for remote command-line access. Like in FTP all data, including credentials are transmitted as plain text. In an OWF, where remote access is critical due to its remote environment, exposed Telnet services could give attackers administrative access to critical components such as turbine controllers or networking equipment. Exploitation could result in complete system takeover, disruption of monitoring systems, or more attacks which target SCADA systems [3].

2.4 The use of Digital Twins

Digital twin (DT) technology in offshore wind farms has enabled organizations to build virtual replicas of turbines, substations, and farm systems that are synchronized with real-world data and provide insight into design, operation, and maintenance. In turn the use of this technology provides benefits from lower costs to improved reliability and resilience in remote offshore environments[18].

A major use is predictive maintenance. DTs combine sensor data, like as vibration, temperature, and torque, with models to predict failures in components like bearings, blades, or gearboxes within turbines. This allows repairs to be scheduled proactively, avoiding unplanned downtime and preventing cascading failures. Many predictive DT platforms rely on OPC UA (Open Platform Communications Unified Architecture), a secure industrial communication standard designed for interoperability, to exchange data with turbine controllers, while still interfacing with legacy SCADA systems [10].

In addition, DTs support fault diagnosis and condition monitoring. They provide real time visualization of turbine health. This capability enables early detection of anomalies, reducing the need for expensive on site inspections and improving safety. Reviews emphasize that DT based monitoring can significantly extend the lifetime of turbines[14].

Despite these benefits, DTs also inherit cybersecurity risks from the OT systems they replicate. OWFs depend on SCADA networks and industrial controllers that still use legacy or insecure protocols. For example, Modbus/TCP is common in turbine telemetry, DNP3 in substation relays, Siemens S7 in turbine PLCs, and FTP/Telnet in maintenance file transfer and remote access. When DTs replicate these systems they inherit the same security concerns [8]. OPC-UA, used in modern DTs, addresses many of these gaps, but its coexistence with legacy protocols in real installations means misconfigurations can still leave systems exposed. Documented assessments confirm that such insecure services have been identified in scans of critical energy infrastructure, including the wind sector [3].

2.5 Research Process

The research process for the information in the state of the art / literature review for this project was conducted using the academic/scientific paper research tools IEEE / Arxiv / Research Gate / Web of Science, using the keywords "Offshore Wind Farms" / "Cybersecurity" / "Digital Twins". Also Microsoft Bing was used as a search engine along with its Microsoft CoPilot AI Companion in order to also find relevant sources and define words/concepts found accross this research.

3 Project Proposal

In a paper which conducts an analysis of cyber security scenarios in wind farm SCADA systems[15], one of the mitigation controls to reduce the attack surface of Wind Farm SCADA systems is to disable the use of insecure protocols. We propose to use this as a starting point in a study which targets OWF devices with services found in the public internet running these protocols. We aim to collect data from these devices' exposed services and :

1. Identify the role of the device in the Offshore Wind Farm scene
2. Identify potential digital twin devices
3. Observe how the data of these devices' services changes over time
4. Explore CVEs relate to the devices' services

We chose to focus on European OWF projects due to convenience as data about them can be found in the EMODnet Map Viewer. Table 1 explains the choice of the protocols we intend to look for in devices which run on offshore wind farms and the reasoning behind their choice.

Protocol	**Justification**	**Role**	**Sources**
Siemens S7	Known for vulnerabilities due to lack of encryption and authentication.	Used in S7 PLCs which are used in turbines.	[19], [11]
DNP3	No authentication or encryption.	Used in PLCs in turbines and substations.	[11]
FTP	All data is transmitted in plain text.	Transferring logs, performance data, or firmware updates.	[15], [16]
Modbus	No authentication or encryption.	Used in SCADA systems with PLCs and RTUs.	[11]
Telnet	Transmits data in plain text, lacks encryption and authentication.	Legacy protocol used for device remote access and diagnostics.	[15]

Table 1: Insecure protocols and their justification.

In order to ensure that the data collected has a high chance of relating to these Offshore Wind Farm projects, we will be using country and organization information when looking at different devices and cross-referencing the information with the data we have collected about the Offshore Wind Farms in Europe.

To collect service and device data we will be using MODAT Magnify which uses AI to collect ścansóf exposed services, as well as extract and organize their data. Taking advantage of the MODAT Magnify's online web tool as well as its API can help us find devices through queries which look for specific device and service information such as location and organizations involved.

Combining MODAT Magnify with python can help us develop a CLI (Command Line Interface) to help us automate processes during the investigation and make it easier for other researchers to replicate or extend this investigation.

The contribution of this project to the field includes:

- a dataset of European OWFs
- a dataset of services running DNP3, FTP, MODBUS, SIMENS S7 and TELNET which belong to OWF devices
- a method for classification of potential roles of devices in OWFss based on service data
- a method of classification for identifying potential digital twins of devices in OWFs
- observations of how service data changed over one month
- an exploration of the Common Vulnerabilities and Exposures (CVEs) associated with the service data we collected
- A python CLI as well as python and bash scripts which automate the processes involved in delivering the contributions above.

This is a meaningful contribution for researchers as it combines the topics of cyber security as well as digital twin identification. Given the scale of investments into offshore wind farm technology in Europe it is important to address the issues caused by the convergence of OT and IT as mentioned in many papers.

4 Methodology

4.1 Tools

In this project we will be using the following tools for distinct roles which can be seen in the table below.

Tool	Feature	Use Case
MODAT Magnify	API and Web Tool	Collect Data of exposed services, their hosts, and their historic data
python	libs: pandas, requests, seaborn, haversine, click	Retrieve data using requests library, process / filter data using pandas, visualize data using seaborn, to create a cli for everything
Github	Repo	to store the code and data I used
IPInfo	Legacy API	Retrieve location info of device using IP
Overleaf Latex	Web Tool	Writing thesis
VSCode	Code Editor	Write code scripts and project CLI
LLMs	-	Discussed in Use of AI section

Table 2: Tools Used for the Project and their use case

4.2 Stages

Since we want to make the study reproducible and potentially extendable by other researchers, all data and scripts will be available in a public GitHub Repository. Stages 3 to 5, will be implemented in a python CLI using click. All data will be stored as CSV files.

- Stage 1 : Choosing Devices
- Stage 2 : Collecting Device Data
- Stage 3 : Classifying Roles of Devices in Offshore Wind Farms
- Stage 4 : Classifying Potential Digital Twins

- Stage 5 : Identifying Changes in Devices and Services Over Time & Collecting CVEs found
- Stage 6 : Writing the thesis

4.2.1 Stage 1 : Choosing Devices

For this project we will be looking for OWF devices in Europe running the 5 insecure protocols we discussed: DNP3, FTP, Modbus, Siemens S7, Telnet. In an attempt to choose plausible OWF devices, we will collect data of all offshore wind farm projects in Europe and the organizations involved in those projects.

Using this information we will undergo a testing phase where we test out queries on MODAT Magnify and record the number of results we get. The queries will be using specific countries for location, specific protocols being and run on specific ports, and the presence of relevant organizations.

We then choose a set of queries whose number of results is manageable, up to 25 results, in case we need to conduct manual checks, and we will retrieve the ip address of those devices found. After that we will retrieve specific location information for each IP and cross reference it with the data of OWFs in Europe, with the goal of finding the 3 closest OWFs of each device as well as the distances to them.

Based on this information we will aim to choose a distance to use as filter such that we have around 10 devices to work with per protocol, for us to use for the remainder of this project. This will include visualization of distance distributions and trial and error in order to obtain the number of results we want: 10 devices per protocol. This should give us 50 devices to use.

Given our filtered data, we use the devices' ips to generate a set of queries to use to retrieve service data for this project. These queries will target for the use of the 5 insecure or legacy protocols we have used so far (DNP3, FTP, Modbus, Siemens S7, Telnet), as well as http, as services running http might present more relevant information for Stage 3 and 4.

4.2.2 Stage 2 : Collecting Device Data

In this stage we will be retrieving service data as well as historical service data (scans of services up to one month old) from MODAT Magnify using the queries generated in the last part of Stage 1. This data will be used for Stages 3,4 and 5, which involve classifying devices by roles, identifying potential digital twins, observing changes in data over time and exploring the CVEs associated with the services we collected data from.

4.2.3 Stage 3 : Classifying Roles of Devices

The next phase of the project will be to classify devices based on service data into potential roles they might have in OWFs. The roles will be based on the network hierarchy architecture intended to be used in OWFs, as well as contain a fallback role for devices whose services' data doesn't provide a clear indication as being part of a particular role[2]. We will also add a final filter, case by mistake we included devices which are clearly unrelated to OWFs. The roles we will be using are:

- Turbine
- Substation
- Control Center
- Other
- Not part of Offshore Wind Farms

We will be using fields in Table 3 to created a condition based ticking system to allocate roles. The general idea is to create a criteria for each role which includes a set of objective conditions such as finding a specific keyword in a device's services' data, and based on the number of conditions met per role criterion, a role is attributed, given that at least 2 conditions are met. Other will be a fallback role, as hinted before, and will be used for devices who don't meet any criteria. The Not Part of OWF criteria will work a bit differently. This criteria's conditions will instantly deem a device as Not Part of OWF, if any hard condition is met, such as the presence of an unrelated service tag, or flag as needing review, for conditions like having a general web tech stack.

Information	**Fields**
Organizations	`asn_org`, `service_tls_issuer_organization`, `service_tls_subject_organization`, `service_tls_issuer_organizational_unit`, `service_tls_subject_organization_organizational_unit`
Hostnames	`fqdns`
Keywords Present	`service_banner`, `service_http_title`, `service_http_body`
Technologies	`service_fingerprints_service_product`, `service_fingerprints_technologies`
Tags	`service_fingerprints_tags`

Table 3: Fields used to create conditions for role classification system

4.2.4 Stage 4 : Classifying Potential Digital Twins

During this stage, we will attempt to identify potential DT devices within the devices which will be attributed roles within OWFs. We will be using a similar classification system with the same principles as the one described for attributing roles but in this case devices will either be labelled as potential DTs or not.

4.2.5 Stage 5 : Identifying Changes in Device Data over time & Collecting CVEs found

The final stage of this project is to observe any data changes to devices' services' data over the period of time we collected data from as well as exploring CVEs found in service data. This data will then be organized by fields changing over time per protocol, offshore wind farm role and in potential digital twins. Some data changes we are not going to deem as relevant such as changes from empty fields to fields having data, banner changes to do with "Date" section(as these are meant to include the different timestamps of when the service was scanned) and hash fields (as they reflect changes in other such as http headers and bodies). This leaves us with observing changes in organization fields, http fields (excluding hash fields), protocols, banners (excluding 'Date' section), tls certificate fields(excluding hash fields), technologies and products.

For the exploration of the CVEs found in our service data, we will extract it and have it in a unique data set and attempt to describe the CVEs associated with different types of devices in offshore wind farms giving explaining the most recent CVEs for the distinct services' protocols they are found in.

4.3 Planning

Week	**Stage**	**Milestones**
1	1	Collect Offshore Wind Farm data and make sets of queries to test using host search
2	1	Collect host search data for queries, get location information for results, cross-reference information to choose set of devices
3	2	Make queries for specific devices, relevant services + HTTP (for more info), collect service search data, collect history data for all services
4	3	Create and apply classification system to service data assigning roles: Turbine, Substation, Control Center, or Not Offshore Wind Farm
5	4	For all devices with roles other than Not Offshore Wind Farm, create and apply classification system to assign labels of Potential Digital Twin or Not Digital Twin
6	5	For all devices with roles other than Not Offshore Wind Farm, identify changes in history data
7	6	Introduction, Literature Review, Proposal, Methodology
8	6	Stages, Conclusions, Evaluation, Use of AI

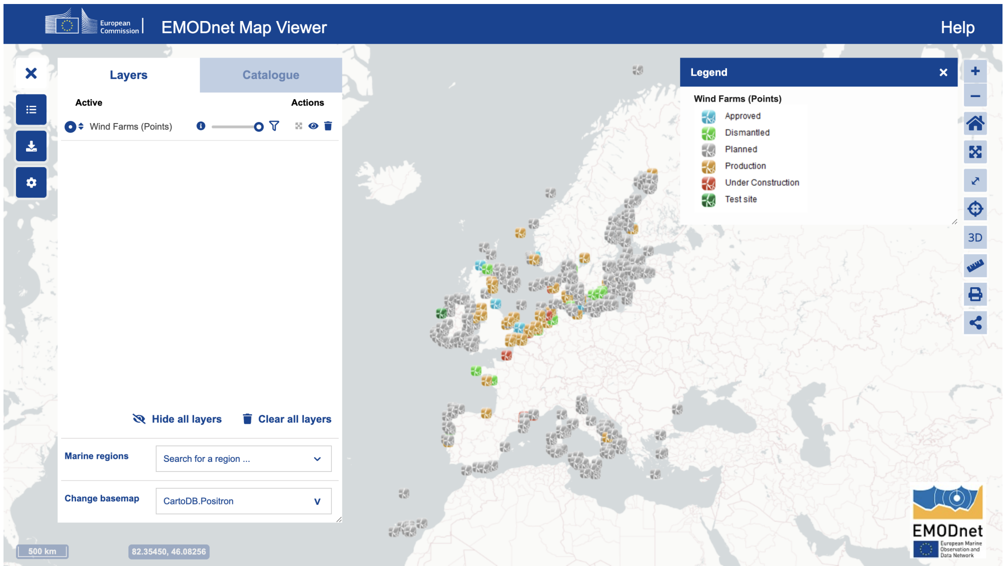
Table 4: Stages and Milestones of this Project

5 Results & Analysis

As mentioned, in order to support the reuse and extension of this study we provide all data and tools used in a Github repository. The repository has a README.md in its root explaining the role of each directory. In summary, scripts and data used in Stage 1 can be found under the **choosing_devices/** and **europe_windfarms/** directories, and the rest of the stages involved the implementation of a CLI found in the directory **project_cli** (which includes .md files on how to use it). The **data/** directory is where the **project_cli** stores data.

5.1 Stage 1 : Choosing Devices

We used EMODnet Map Viewer to extract data of all Offshore Wind Farm projects in Europe as a json file.



Change basemap CartoDB.Positron

Figure 1: EMODnet Map of Offshore Wind Farms in Europe.

After filtering the data, using a python script with pandas, to include only Offshore Wind Farms with more than one turbine and whose status is Production implying that they were active Offshore Wind Farms (**europe_windfarms\notebook.ipynb** in Repository). In summary we had a data set of 60 offshore wind farms.

Using our Offshore Wind Farm data set, we then used LLM models (with their research modes) in order to compile lists of vendor and operator organizations involved in each project. The models used are mentioned in the Use of AI section and the prompts can be found in the project repository's **europe_windfarms \notebook.ipynb** This will information will then be used in order to make sets of queries using the organization, country of offshore wind farms and the protocols we are looking for (Siemens S7, Modbus, Telnet, FTP, DNP3).

Given the information of offshore wind farms, we created a series of 150 queries with the aid of LLMs to used in the host search feature of MODAT Magnify web tool, and registered the number of results each query had. All queries looked for organization names in the org related fields of MODAT Magnify's service and device data as well as web.html and banner data of services and devices. The form of each query was as follows:

(org ”org_name” or web.html ”org_name” or banner ”org_name”) and service=protocol_name and port=protocol_port and country=country_2_iso_code and last_seen> 2024-07-24

The last_seen field was to ensure that the records we were using were recent and hence we could retrieve the history records in the next stage since our MODAT Magnify membership allows us to retrieve history data of service upto one month old. Below are a couple of example of queries we used and the number of results they got.

Query	Number of results
org~"Siemens" and service=dnp3 and port=20000 and country=DE and last_seen>2024-07-24	1
(org~"Windmw" or web.html~"Windmw" or banner~"Windmw") and service=ftp and port=21 and country=DE and last_seen>2024-07-24	2
(org~"eneco" or web.html~"eneco" or banner~"eneco") and service=modbus and port=502 and country=DE and last_seen>2024-07-24	1
org~"Siemens" and service=siemens and port=102 and country=DE and last_seen>2024-07-24	25
(org~"eneco" or web.html~"eneco" or banner~"eneco") and service=telnet and port=23 and country=DE and last_seen>2024-07-24	1

Table 5: Sample of queries used for choosing devices

After recording all the queries and the number of results each had, we decided to choose a subset of queries whose results were below 25 records as this was a number that we could work with manually if needed unlike results much larger. The subset of queries was then all put into a table where we checked if the same query with no port had more results for us to use as more services could be running the same protocol in different ports. Depending on whether which query had more results, we then used a couple of simple bash scripts in order (found in the Repository under **queries\bash_scripts**) to use the MODAT Magnify API host_search endpoint , in order to retrieve the ip address of all results and then retrieve their location information using the IPInfo Legacy API. This data was then added to the relevant query data sets. Using python scripts (**queries\check_distance.py** in the Repository) with the haversine and pandas library, we cross-referenced each ip's location with the location of each offshore wind farm in order to make a new data set with the 3 closes offshore wind farm per device. In order to properly filter the data with what was relevant, we then used the seaborn library in a python script (**queries\choosing_devices.py** in the repo) to visualize the distribution of distances (between offshore wind farms and devices) in order to find what would be an appropriate filter, based on location, to apply to our devices so that we have around 10 results per protocol query set (queries searching for the 5 individual protocols we have discussed).

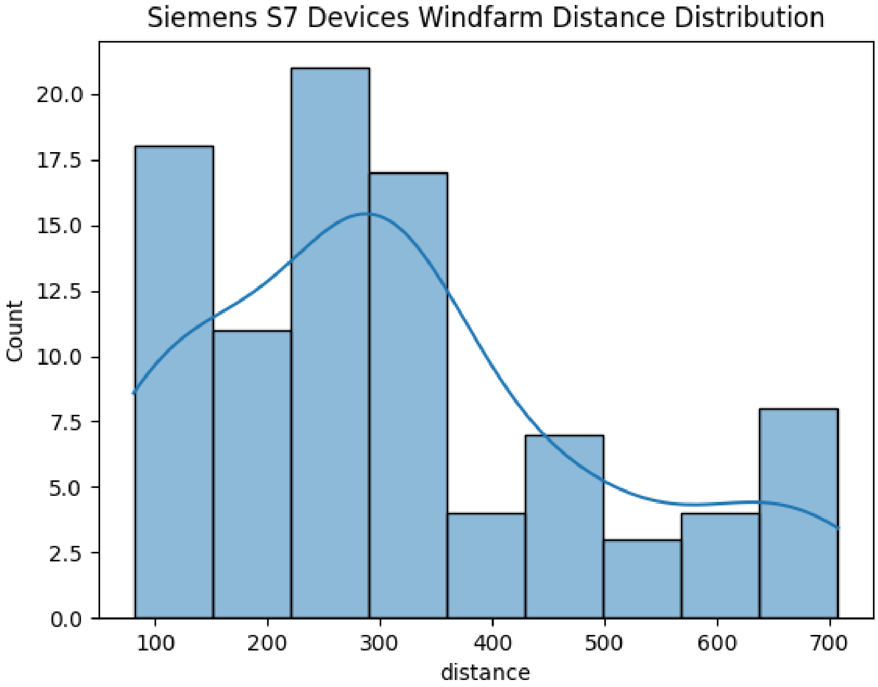


Figure 2: Sample Histogram of distances of devices found running Siemens S7 protocols in one of our sets of queries' results (**gpt_queries_2_relevant_check_windfarms.csv**). In this case we chose to apply the filter < 100 km which gave us 14 records

The aftermath led us to conduct a trial and error basis where we start by testing applying a filter of distances < 100 km, if there is little or no results we increment the distance filter by 50 km until 200 km, where if the number of values doesn't change, we subset of data with results after 100 km. If there is too much data, we try and use a filter of 50 km distance. Using the **filtering.py** python script we apply these filters.

Protocol	Distance filter of devices from Offshore Wind Farms (km)	Number of Results
DNP3	200	1
FTP	50	13
Modbus	100	6
Siemens S7	100	14
Telnet	100	1

Table 6: Sample Distance Filter and Number of Results per Protocol for one of the sets of queries we ran (**chosen_queries_windfarms.csv**)

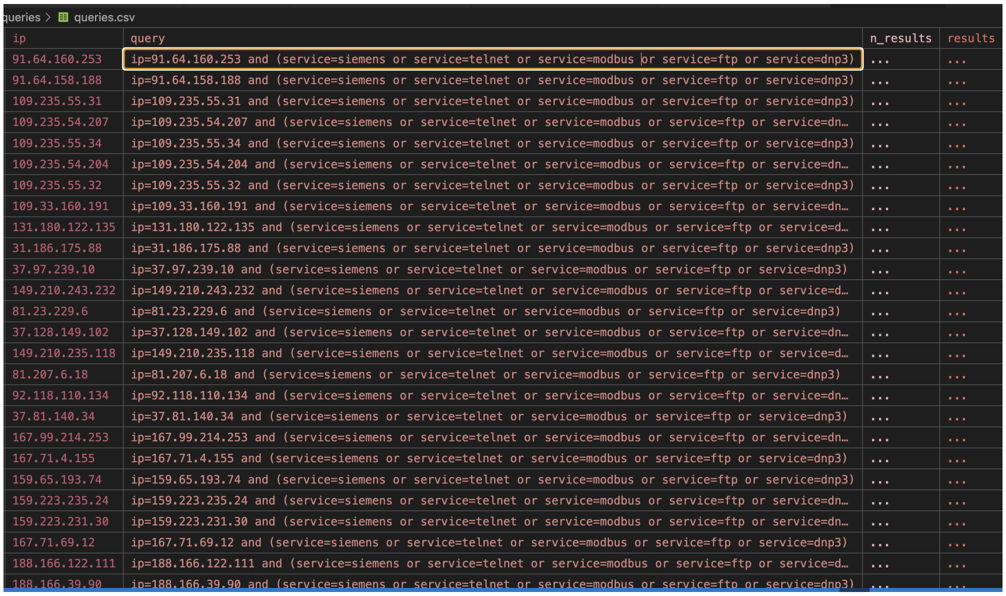
We then generate a csv with all the unique ip values and the protocols found under them (using the **queries\generate_queries.py** in the Repository), so we can then make specific device queries (i.e. use the ip) to run service searches in the next stage of the project. These queries look for all 5 protocols we used to find the devices as well as http protocols within each individual ip.

Protocol	Number of Devices
DNP3	1
FTP	24
Modbus	3
Siemens S7	21
Telnet	1
Total	49

Table 7: Summary of devices chosen and insecure or legacy protocols found running in them (one device found with Modbus and Siemens S7)

5.2 Stage 2 : Collecting Device Data

Applying the **generate_queries.py** to the csv files of individual ip addresses and the protocols found running on them, made a new csv with the columns 'ip', 'query', 'n_results' and 'results'. The columns including a results column had placeholder values to be filled with the results of the service searches we conducted next. We ended up with 2 csv files with 50 queries. One of the files is to store the queries looking for the 5 protocols we wanted and the other is used for queries which simply looks for http protocols in order to use for more information later.

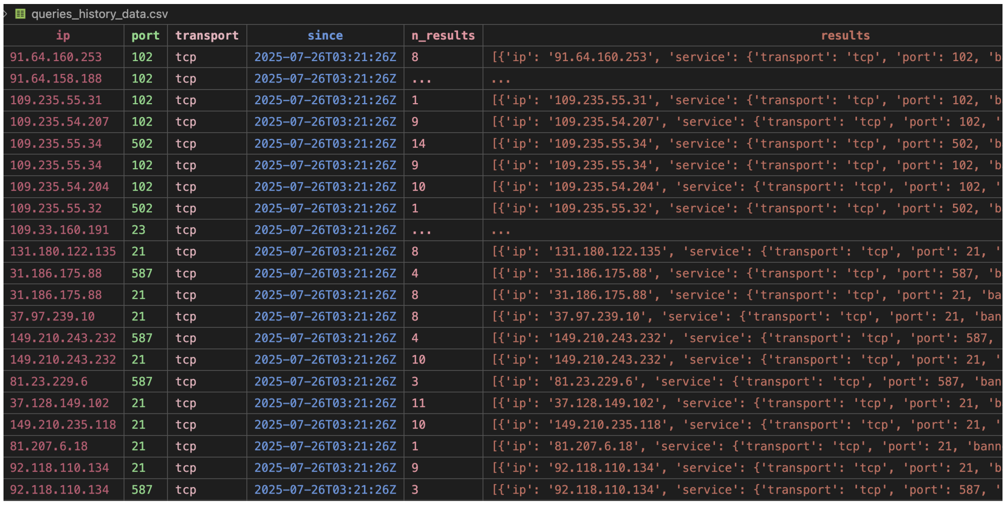


queries > queries.csv

ip	query	n_results	results
91.64.160.253	ip=91.64.160.253 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
91.64.158.188	ip=91.64.158.188 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
109.235.55.31	ip=109.235.55.31 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
109.235.54.207	ip=109.235.54.207 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
109.235.55.34	ip=109.235.55.34 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
109.235.54.204	ip=109.235.54.204 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
109.235.55.32	ip=109.235.55.32 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
109.33.160.191	ip=109.33.160.191 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
131.180.122.135	ip=131.180.122.135 and (service=siemens or service=telnet or service=modbus or service=ftp or service=d…	...	...
31.186.175.88	ip=31.186.175.88 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
37.97.239.10	ip=37.97.239.10 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
149.210.243.232	ip=149.210.243.232 and (service=siemens or service=telnet or service=modbus or service=ftp or service=d…	...	...
81.23.229.6	ip=81.23.229.6 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
37.128.149.102	ip=37.128.149.102 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
149.210.235.118	ip=149.210.235.118 and (service=siemens or service=telnet or service=modbus or service=ftp or service=d…	...	...
81.207.6.18	ip=81.207.6.18 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
92.118.110.134	ip=92.118.110.134 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
37.81.140.34	ip=37.81.140.34 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
167.99.214.253	ip=167.99.214.253 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
167.71.4.155	ip=167.71.4.155 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
159.65.193.74	ip=159.65.193.74 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
159.223.235.24	ip=159.223.235.24 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
159.223.231.30	ip=159.223.231.30 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dn…	...	...
167.71.69.12	ip=167.71.69.12 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...
188.166.122.111	ip=188.166.122.111 and (service=siemens or service=telnet or service=modbus or service=ftp or service=d…	...	...
188.166.39.90	ip=188.166.39.90 and (service=siemens or service=telnet or service=modbus or service=ftp or service=dnp3)	...	...

Figure 3: CSV created with queries for individual devices running insecure protocols (siemens s7, modbus, telnet, ftp, dnp3)

Using a CLI we generated using python with the library click, we ran a command which uses the csv files, runs the queries using the MODAT Magnify API's service search endpoint and stores their result data. Then we ran another command which takes a csv with the results and create a new csv which aggregates the response data of all queries' results. The extracted service search response data csv was then used by another cli command in order to create a csv with info for history queries. Using the history queries we used another of the cli commands which retrieves the information of each service history and stores it.



queries_history_data.csv

ip	port	transport	since	n_results	results
91.64.160.253	102	tcp	2025-07-26T03:21:26Z	8	[{'ip': '91.64.160.253', 'service': {'transport': 'tcp', 'port': 102, 'b
91.64.158.188	102	tcp	2025-07-26T03:21:26Z	...	...
109.235.55.31	102	tcp	2025-07-26T03:21:26Z	1	[{'ip': '109.235.55.31', 'service': {'transport': 'tcp', 'port': 102, 'b
109.235.54.207	102	tcp	2025-07-26T03:21:26Z	9	[{'ip': '109.235.54.207', 'service': {'transport': 'tcp', 'port': 102, '
109.235.55.34	502	tcp	2025-07-26T03:21:26Z	14	[{'ip': '109.235.55.34', 'service': {'transport': 'tcp', 'port': 502, 'b
109.235.55.34	102	tcp	2025-07-26T03:21:26Z	9	[{'ip': '109.235.55.34', 'service': {'transport': 'tcp', 'port': 102, 'b
109.235.54.204	102	tcp	2025-07-26T03:21:26Z	10	[{'ip': '109.235.54.204', 'service': {'transport': 'tcp', 'port': 102, '
109.235.55.32	502	tcp	2025-07-26T03:21:26Z	1	[{'ip': '109.235.55.32', 'service': {'transport': 'tcp', 'port': 502, 'b
109.33.160.191	23	tcp	2025-07-26T03:21:26Z	...	...
131.180.122.135	21	tcp	2025-07-26T03:21:26Z	8	[{'ip': '131.180.122.135', 'service': {'transport': 'tcp', 'port': 21, '
31.186.175.88	587	tcp	2025-07-26T03:21:26Z	4	[{'ip': '31.186.175.88', 'service': {'transport': 'tcp', 'port': 587, 'b
31.186.175.88	21	tcp	2025-07-26T03:21:26Z	8	[{'ip': '31.186.175.88', 'service': {'transport': 'tcp', 'port': 21, 'ba
37.97.239.10	21	tcp	2025-07-26T03:21:26Z	8	[{'ip': '37.97.239.10', 'service': {'transport': 'tcp', 'port': 21, 'ban
149.210.243.232	587	tcp	2025-07-26T03:21:26Z	4	[{'ip': '149.210.243.232', 'service': {'transport': 'tcp', 'port': 587,
149.210.243.232	21	tcp	2025-07-26T03:21:26Z	10	[{'ip': '149.210.243.232', 'service': {'transport': 'tcp', 'port': 21, '
81.23.229.6	587	tcp	2025-07-26T03:21:26Z	3	[{'ip': '81.23.229.6', 'service': {'transport': 'tcp', 'port': 587, 'ban
37.128.149.102	21	tcp	2025-07-26T03:21:26Z	11	[{'ip': '37.128.149.102', 'service': {'transport': 'tcp', 'port': 21, 'b
149.210.235.118	21	tcp	2025-07-26T03:21:26Z	10	[{'ip': '149.210.235.118', 'service': {'transport': 'tcp', 'port': 21, '
81.207.6.18	21	tcp	2025-07-26T03:21:26Z	1	[{'ip': '81.207.6.18', 'service': {'transport': 'tcp', 'port': 21, 'bann
92.118.110.134	21	tcp	2025-07-26T03:21:26Z	9	[{'ip': '92.118.110.134', 'service': {'transport': 'tcp', 'port': 21, 'b
92.118.110.134	587	tcp	2025-07-26T03:21:26Z	3	[{'ip': '92.118.110.134', 'service': {'transport': 'tcp', 'port': 587, '

Figure 4: Histories csv already filled with all service history info of services running insecure protocols.

Protocol	Number of Results
DNP3	1
FTP	41
HTTP	159
Modbus	3
Siemens S7	21
Telnet	16

Table 8: Summary of Device services collected per protocol

Protocol	Number of Results	Average number of records per service per protocol (i.e. histories) (3 s.f.)
DNP3	1	1
FTP	176	4.2
HTTP	732	4.6
Modbus	16	5.3
Siemens S7	99	4.7
Telnet	25	1.6
Unknown	2	-

Table 9: Summary of device services per protocols, including histories

5.3 Stage 3 : Classifying Roles of Devices in Offshore Wind Farms

5.3.1 Classification and Roles

In this project we have decided to classify devices based with the following categories:

- Turbine
- Substation
- Control Center
- Other
- Not Part of Offshore Wind Farm

The types of information we will be using as well as the type of classification were mentioned in Table 3. The basic idea is to create a tick based system to attribute roles to offshore wind farm devices. Each role has a criteria with a number of conditions, for each condition met a tick is attributed to that role's ticks and the role with the most points, given they are larger than 2, for each device is the role we use. The only 2 roles which work differently are "Not Part of Offshore Wind Farm" and "Other". Other is a fallback condition where if no role can be attributed (i.e. have more than 2 points) then the device is considered "Other". Given that any condition of "Not Part of Offshore Wind Farm is met, then the device is instantly classified as "Not Part of Offshore Wind Farm". Table 10 contains with the number of conditions per role. The criterion and conditions used in them were created with the aid of LLMs which helped brainstorm possibilities as well as come up with more keywords, organizations, technologies, for the conditions using the sources in the bibliography. Table 11 is the criteria of the Turbine role which includes the type of information used (which tells us what fields they look at again Table 3) and provides insight on their choice. All tables of device role and criteria can be found in the appendix(Table 27, Table 27, Table 28, Table 29).

Role	**Number of Conditions**
Turbine	8
Substation	8
Control Center	7
Not Part of OWF	3

Table 10: Number of classification conditions per role (excluding *Other* as it is a fallback role).

Condition	**Criteria (match any)**	**Information (fields)**	**Justification**
T1 Turbine semantics?	nacelle, yaw, pitch, WTG, turbine, tower, hub, blade;	Keywords: `service_banner`, `service_http_title`, `service_http_body`	IEC 61400-25 models turbine subsystems via logical nodes → strong turbine context.
T2 "IEC 61400-25" seen?	Contains "61400-25" / "IEC 61400-25"	Keywords (same)	Widely used in wind operations/SCADA stacks.
T3 Turbine nets?	TAN, FAN, CAN	Keywords (same)	Common wind plant topology shorthand; supportive heuristic.
T4 PLC/S7 + turbine word?	S7/SIMATIC/S7-1200/1500/300/400 *and* a T1 word (e.g., nacelle/yaw/pitch)	Tech: `service_fingerprints_service_product`, `service_fingerprints_technologies`; Keywords	PLCs in tower/nacelle are typical; pairing PLC with turbine terms increases precision.
T5 OPC-UA at boundary?	opc ua/opc-ua/opcua + (SCADA | OSS | DMZ | control center | CAN)	Keywords; Tech	OPC-UA often links turbine/substation layers upward to SCADA/control center.
T6 OT/SCADA tags?	SCADA, ICS, OT, Smart Energy, Power Plant	Tags: `service_fingerprints_tags`	Dataset OT/energy tags support turbine role.
T7 OEM/operator org + turbine word?	Org has e.g., Vestas, Siemens Gamesa *and* a T1 word	Organizations (ASN/TLS); Keywords	Org + turbine semantics raises confidence; avoids brand-only hits.
T8 Turbine FQDN hint?	Regex: `\b(turbine|wtg|nacelle)\b`	Hostnames: `fqdns`	Role-encoded hostnames are supportive; reviewable.

Table 11: Turbine role—conditions, criteria, fields, and justification.

This system was then implemented in the CLI of this project (**project_cli\** in the repository) in a way that the criteria and the data it uses have a modular design so it can be easily extended or changed. Data files which include keywords or organization names are stored by Criteria and by role under the **classification_criteria_roles/** directory in my GitHub repository. Each role has a set of directories, one for each condition of the role's criteria, where the information used for condition is stored. The actual logic is implemented in the **classify_roles.py** and **roles_class_helpers.py**.

Role	Count
Turbine	1
Substation	13
Control Center	2
Other	26
Not part of Offshore Wind Farms	7

Table 12: Summary of devices classified by roles in offshore wind farms

5.3.2 Sample Classification Turbine

Attribute	Value
ip	167.71.4.155
predicted_role	Turbine
ticks_Turbine	2
ticks_Substation	1
ticks_ControlCenter	2
ticks_NotPartOfOWF	0
fired_Turbine	T1_keywords, T3_tan_tbn_fan
fired_Substation	S3_ied_rtu_bay
fired_ControlCenter	C4_enterprise_platforms, C5_remote_access
fired_NotPartOfOWF	
hostnames	167.71.4.155

Table 13: Sample Classification of Turbine Role for 167.71.4.155

5.3.3 Sample Classification Substation

Attribute	**Value**
ip	190.105.195.4
predicted_role	Substation
ticks_Turbine	0
ticks_Substation	3
ticks_ControlCenter	1
ticks_NotPartOfOWF	0
fired_Turbine	
fired_Substation	S3_ied_rtu_bay, S6_tso_org, S7_tags
fired_ControlCenter	C6_operator_org
fired_NotPartOfOWF	
hostnames	

Table 14: Sample Classification of Substation Role for 190.105.195.4

Condition 3 of the Substation criteria fired as ”sel” was found in banners. Condition 6 of the Substation criteria was fired due to ”Terna” being found in organization names. Condition 7 of the Substation criteria was fired due to the presence of the OT tag.

5.3.4 Sample Classification Control Center

Attribute	**Value**
ip	109.33.160.191
predicted_role	Control Center
ticks_Turbine	1
ticks_Substation	3
ticks_ControlCenter	4
ticks_NotPartOfOWF	0
fired_Turbine	T3_tan_tbn_fan
fired_Substation	S1_iec61850_mms_goose, S3_ied_rtu_bay, S6_tso_org
fired_ControlCenter	C1_cc_labels, C4_enterprise_platforms, C5_remote_access, C6_operator_org
fired_NotPartOfOWF	
hostnames	

Table 15: Sample Classification of Control Center Role for 109.33.160.191

Condition 3 for the Turbine criteria fired as ”tan” was found in the http body of the services running http. Condition 1 of the Substation criteria fired due to the presence of the keyword ”goose” in the http service running on port 81. Condition 3 of the Substation criteria was fired due to the presence of ”ied”, ”rtu”, ”sel” in the http body and banners of http services.

5.3.5 Sample Classification Not Part of OWF

Attribute	**Value**
ip	149.210.243.232
predicted_role	Not Part of OWF
ticks_Turbine	2
ticks_Substation	3
ticks_ControlCenter	1
ticks_NotPartOfOWF	1
fired_Turbine	T1_keywords, T3_tan_tbn_fan
fired_Substation	S1_iec61850_mms_goose, S3_ied_rtu_bay, S4_poc_pcc
fired_ControlCenter	C4_enterprise_platforms
fired_NotPartOfOWF	N3_generic_it_ui, manual_override
hostnames	cablejuice.tv cpanel.cablejuice.tv cpcalendars.cablejuice.tv cpcontacts.cablejuice.tv mail.cablejuice.tv webdisk.cablejuice.tv webmail.cablejuice.tv www.cablejuice.tv

Table 16: Sample Classification of Not Part of OWF Role for 149.210.243.232

In the case of 149.210.243.232, although 3 conditions were met for the Substation role, the generic IT UT condition was fired by the Not Part of OWF role. This implied a manual check of the device based on the use of generic web tools and after that and obvious indications from the hostnames of the device, which implied it wasn't related to OWFs, this device was classified as Not Part of OWF.

5.3.6 Sample Classification Other

Attribute	**Value**
ip	159.65.193.74
predicted_role	Other
ticks_Turbine	1
ticks_Substation	1
ticks_ControlCenter	1
ticks_NotPartOfOWF	0
fired_Turbine	T1_keywords
fired_Substation	S3_ied_rtu_bay
fired_ControlCenter	C4_enterprise_platforms
fired_NotPartOfOWF	
hostnames	

Table 17: Sample Classification of Other Role for 159.65.193.74

5.4 Stage 4: Classifying Potential Digital Twins

Based on the results of the subsection 5.3 then created a similar point/tick-based point system in order to classify the devices with roles part of offshore wind farms as being potential digital twins or not. Table 19 shows and justifies the criteria used. There are 3 types of conditions and they attribute different points to a device: Hard (3 points), Medium (2 points) and Soft (1 point). We have developed 2 Hard conditions, 4 Medium Conditions and 6 Soft Conditions. Given a device is attributed at least 2 points (i.e. one Medium Condition), it is labelled as a potential digital twin. Once again, the way we reached to this criteria was by using LLMs to find keywords, organizations and technologies within the sources in the bibliography.

Table 18: Single-role classifier "Potential Digital Twin": conditions, criteria, fields, and justification.

Condition	**Criteria (match any)**	**Information (fields)**	**Justification**
DT1 Digital-twin terms?	`digital twin`, `digital-twin`, `digital twin platform`, `virtual twin`, `cyber-physical twin`, `twin data`	Keywords: `service_banner`, `service_http_title`, `service_http_body`	Direct textual signal of a DT or DT platform.

Continued on next page

Condition	**Criteria (match any)**	**Information (fields)**	**Justification**
DT2 DT platforms/vendors?	Azure Digital Twins; Siemens MindSphere; PTC ThingWorx; Ansys Twin Builder; GE Predix; IBM Maximo APM; AVEVA PI / Data Hub	Tech/Products: `service_fingerprints_service_product`, `service_fingerprints_technologies`; Orgs: `asn_org`, TLS org/OU; Keywords (same)	Named DT/industrial-cloud platforms strongly indicate DT deployments.
DT3 Wind modeling tools?	OpenFAST; FAST.Farm; HAWC2; Bladed; Flex5	Tech/Products (same); Keywords (same)	Wind DT pipelines rely on aeroelastic/physics models; presence implies simulation–plant linkage.
DT4 Sim/dev/test wording?	`HIL`, `SIL`, `testbed`, `simulator`, `simulation`, `emulation`, `sandbox`, `training`, `staging`, `preprod`, `demo`	Keywords (same)	DTs are exercised/validated via SIL/HIL and test/staging environments.
DT5 Predictive / PHM / O&M?	`RUL`, remaining useful life; `PHM`; prognostics; anomaly detection; early warning; degradation; condition monitoring; CBM; predictive maintenance; failure prediction	Keywords (same)	DTs support prognostics and maintenance decision-making.
DT6 IIoT / data connectors?	OPC UA/OPC-UA; MQTT (EMQX, Mosquitto, HiveMQ); AMQP; Kafka; RabbitMQ; Azure IoT Hub; AWS IoT Core; IoT Edge; Kinesis; Kepware KEPServerEX; Matrikon OPC Server	Tech/Products; Orgs; Keywords (same)	DTs need telemetry ingest/streaming; these stacks are typical data plumbing.
DT7 Analytics/visualization?	Grafana; Kibana; Jupyter/Lab; Power BI; Tableau; Apache Superset; Plotly/Dash	Tech/Products; Keywords (same)	Visualization of real+virtual signals is common in DT observability.
DT8 3D/interactive viz engines?	Unity/Unity3D; Unreal Engine; Three.js; CesiumJS; Babylon.js; WebGL	Tech/Products; Orgs; Keywords (same)	3D/interactive front-ends frequently surface DT states for operators.
DT9 Cloud/edge framing?	Tags `Cloud`, `Smart Energy`; Orgs Microsoft/AWS/Google; keywords: Azure, AWS, Google Cloud/GCP, edge computing, cloud IoT	Tags: `service_fingerprints_tags`; Orgs (ASN/TLS); Keywords (same)	Many DTs are deployed in cloud/edge contexts; supportive when paired with OT terms.
DT10 DT-ish hostnames?	Subdomains: `dev.`, `test.`, `staging.`, `demo.`, `lab.`, `sandbox.`, `preprod.`, `training.`, `qa.`; tokens: `digitaltwin`, `twin`, `sim`	Hostnames: `fqdns`	Non-production/twin environments often use these naming patterns (soft but useful).
DT11 SCADA/standards terms?	`SCADA`, historian, HMI, OPC; `IEC 61400-25`; `IEC 61850`	Keywords (same)	DTs integrate with plant/SCADA data and standards; these raise operational DT confidence.

Continued on next page

Condition	Criteria (match any)	Information (fields)	Justification
DT12 API docs endpoints?	`swagger`, `openapi`, `redoc`, `api docs`, `api documentation`, `swagger-ui`	Keywords (same)	Public DT/monitoring APIs often expose documentation UIs; supportive evidence.

Table 19 Shows the results we got after applying the classification to the 16 devices who were attributed roles in offshore wind farms in subsection 5.3.

Potential Digital Twin	Count
Yes	5
No	11

Table 19: Conclusion of Offshore Wind Farm Devices and Potential Digital Twins

5.4.1 Sample Classification Potential Digital Twin

Table 20: Digital Twin Detection Result

Field	Value
IP Address	131.180.122.135
DT Score	3
DT Label	PotentialDigitalTwin
DT Fired Conditions	DT4, DT9
Hostnames	[]

5.4.2 Sample Classification Not Potential Digital Twin

Table 21: Sample Not Potential DT Classification for 83.149.118.89

Field	Value
IP Address	83.149.118.89
DT Score	0
DT Label	
DT Fired Conditions	
Hostnames	[]

No conditions were fired, hence no points were attributed. This left us with no information to suspect this device was a DT using our classification system.

5.5 Stage 5: Identifying Changes in Devices and Services Over Time & Collecting CVEs Found

5.5.1 Changes over time

As shown in the table of devices classified by roles Table 12, there are 18 devices which were classified as part of offshore wind farms after filtering the services data only keeping the devices with the roles Turbine, Substation, Control Center. Using this subset of devices' service data, combined with the historic data of those records of those services, we want to identify changes in the fields of individual services. Some changes are meant to be so we do not take them into account, namely banner data has 'Date' fields which are supposed to change as they are from service data collected at different types. Also we disregard changes from empty values to values as we want to find changes in the data presented and hence we only look at the values from the point they have data. Table 22 shows the different ip's from the filtered devices and information about the number of changes.

IP	Role	Potential DT?	Number of Changes	Fields Changed
149.210.235.118	Substation	Yes	28	service_banner, service_http_body, service_http_title, service_tls_issuer_common_name, service_tls_raw, service_tls_serial_number, service_tls_subject_name, service_tls_expires_at, service_tls_valid_from
37.97.239.10	Substation	Yes	23	service_banner, service_http_body
46.17.3.107	Substation	No	6	service_banner
37.81.140.34	Substation	No	2	service_banner, service_http_body
167.71.4.155	Turbine	Yes	3	service_banner, service_http_body, service_http_title

Table 22: Number of Data relevant changes per device.

The changes for 149.210.235.118 regarding tls fields were related to the fact that a service changed certificates. Most of the banner changes were related to the fact that we didn't take into account cookies, hence any new connection with a http service was given a new "Cookie-Set: session=" value. Other http banners had redirect changes where both scans had code 303 with a note to see new location and that location changed.

5.5.2 CVEs Found

There were 486 different CVEs found across 4 devices in 8 different services runnning only 2 different protocols: http and ftp. Only 5 CVEs are Known Exploited Vulnerabilities(KEV)

IP	Port	Protocol	Device Role	CVE Count
37.97.239.10	21	ftp	Substation	5
37.97.239.10	80	http	Substation	135
37.97.239.10	443	http	Substation	143
131.180.122.135	80	http	Control Center	96
149.210.235.118	21	ftp	Substation	5
149.210.235.118	443	http	Substation	100
167.71.4.155	80	http	Turbine	1
167.71.4.155	443	http	Turbine	1

Table 23: CVEs per service

We will dive deeper into finding out about the most recent vulnerabilities for each different service (based on port and protocol) for the ip which had the most CVEs flagged by far: 37.97.239.10.

Port	Protocol	CVE ID	CVE CVSS	Explanation
21	ftp	CVE-2021-40524	7.5	Pure-FTPd versions 1.0.23 through 1.0.49 contain a logic flaw in the max_filesize quota check, allowing users to upload files of unlimited size. This bypass can lead to denial-of-service by exhausting system resources. The vulnerability is rated high severity and can be exploited remotely. The issue was patched in version 1.0.50, and upgrading is essential for systems relying on quota enforcement.
80	http	CVE-2024-25117	9.8	This vulnerability is related to a php svg lib to parse and render SVGs. SVGs often contain text attributes and the font-family attribute failed to properly validate it allowing attackers to inject malicious php archive paths that can be used to load and execute code from these archives.
443	http	CVE-2025-1695	5.3	This vulnerability affects NGINX, when it is configured to run Java applications, its Java plugin allowed specifically crafted requests to trigger infinite loops which led to excessive CPU usage and even limited Denial of Service. The system doesn't crash entirely but performance is degraded leaving apps potentially unresponsive.

Table 24: Most recent CVEs for different services of 37.97.239.10

5.6 Project CLI

For the purpose of producing a reproducible methodology, we have created a python CLI using the click package for this project in order to automate many of the processes namely:

- the retrieval and formatting of data using the MODAT API
- classification of data (roles and identifying potential digital twins)
- identifying relevant changes in historic data

We also ended up adding features for retrieving IP location and cross-referencing with locations of offshore wind farms in Europe, based on the scripts used in subsection 5.1. In order for researchers to potentially extend or adapt the CLI for their own investigations, we decided to use a clean structure where **main.py** simply adds commands to the CLI but all commands themselves are defined in individual files under the **commands/** folder. To make the code base even more readable and maintainable, we decided to store bigger functions in individual files under the **utils/** folder.

The classification systems, which are an essential piece of this project, were designed and implemented in a modular way so that they can also be extended or adapted. Under the **classification_criteria/**, there are individual folders (**roles/** and **dts/**) which store the data used for the criterion's conditions, organized by criterion and by condition. For example, the keywords used in condition 1 for the Turbine criteria of the role classification system is stored under **classification_criteria/roles/ Turbine/1/keywords.csv**.

6 Conclusion

For the conclusions for this project, we refer back to the research questions and answer them. Limitations and Evaluations are discussed in the next section section 7. In terms of deliverables, i.e. the data we collected as well as the cli and scripts which led us to it, they can all be found in the github repository for this project.

6.1 Can we discover plausible offshore wind devices using insecure or legacy protocols in the public Internet?

Using location, service protocol and organization information we were able to test out around 150 queries which can be found under the **choosing_devices/** directory in the project's GitHub Repository, whose aim is to find devices running services which use DNP3, FTP, MODBUS, SIEMENS S7 and TELNET. The data we collected from offshore wind farm projects, which we used in order to integrate the information about location and organizations involved in offshore wind farm projects in Europe in our queries, can be found under the **windfarms_europe/** directory in the project's Github Repository. After filtering queries by the number of results obtained and filtering the results based on distances to closest offshore wind farms, this project was able to discover 49 devices running services with legacy or insecure protocols.

6.2 Can we classify offshore wind devices found in the public Internet into roles based on service metadata?

After querying the 49 distinct devices which we deemed as plausibly related to OWFs in Europe for the chosen legacy or insecure protocols as well as services running http, we developed a classification system to assign roles within OWFs to the devices. These roles included Not Part of OWF and Other as a final filter of devices which had no indications of being part of a specific role and devices with clear indications of not being part of offshore wind farms at all. After applying our classification system we came to the conclusion that only 16 of our devices could be attributed a role in OWFs, 7 were exlcuded as being part of OWFs and 26 were unclassifiable, which led us to label them simply as other.

6.3 Can we identify potential digital twins within offshore wind devices found in the public Internet based on service meta-data?

We were able to create and implement a DT classification system which led us to believe there were 5 of the 16 devices which we labelled with OWF roles to be digital twins.

6.4 What types of service data changes can be found in offshore wind devices found in the public Internet?

In the case of this experiment the data changes we observed were close to none as we tried to avoid changes we knew would happen between scans such as date values in banners but we forgot to take into account session cookies being set. Other changes included changes in links for new location of resources and a TLS certification change due to expiration. This could be due to the limited time frame we observed the changes for.

6.5 Can we find associated CVEs in offshore wind devices found in the public Internet?

In the case of this experiment the data changes we observed were close to none as we tried to avoid changes we knew would happen between scans such as date values in banners but we forgot to take into account session cookies being set. Other changes included changes in links for new location of resources and a TLS certification change due to expiration. This could be due to the limited time frame we observed the changes for.

7 Limitations & Evaluation

7.1 Choosing Devices

Limitations

- European Scope
- Choosing queries with ≤ 25 results from initial test queries
- Number of queries tested

Improvements

- Global Scale
- Choose test queries which generated more results
- Organization data could be further developed
- Test more queries

7.2 Collecting Data

Limitations

- Only collecting data for dnp3, ftp, http, modbus, siemens s7, telnet
- Collect data over a longer period of time

Improvements

- target devices running more insecure or legacy protocols
- Collect data for all services related to devices chosen

7.3 Classification Systems

Limitations

- Number of Roles
- Number of Conditions used per Role
- Number of Keywords, Technologies, Organizations used in Conditions
- Regex can't deduce all unrelated hostnames (a devices' hostnames contained 'advogados' which is Portuguese for lawyers which is clearly not applicable to offshore wind farms)
- looking for partial words like "ied" is very misleading

Improvements

- Take honeypots in consideration
- More data filters could be applied (more key words, technologies, organizations per role)
- integrate LLM/ML models to classify ambiguous data such as html content, hostnames, etc.

8 Use of AI

This project has references of the use of LLMs and the following table discussed which ones and their use.

LLM Model	Tool	Use Case
ChatGPT 5	ChatGPT web tool and Microsoft Bing Copilot	Extract info from large datasets to help build classification systems; research operators and vendors involved in Offshore Wind Farm projects. ; Extract information from large sources ; Research and finding sources ; Latex Formatting issues
ChatGPT 4.1	GitHub Copilot Chat extension for VSCode	Code debugging.
DeepSeek 3	DeepSeek web tool	Research operators and vendors involved in Offshore Wind Farm projects.
Perplexity AI	Perplexity web tool (uses best model per prompt)	Research operators and vendors involved in Offshore Wind Farm projects.

Table 25: AI tools used in this project

References

[1] Global Offshore Wind Report 2023. Annual report, April 2024.

[2] Mohamed A. Ahmed and Young-Chon Kim. Hierarchical communication network architectures for offshore wind power farms. *Energies*, 7(5):3420–3437, 2014.

[3] Cisco Systems, Inc. Cisco Solution for Renewable Energy: Offshore Wind Farm 1.0 Design Guide. Technical report, Cisco Systems, Inc., May 2023. Design Guide.

[4] Megan Egan. A Retrospective on 2022 Cyber Incidents in the Wind Energy Sector and Building Future Cyber Resilience. Graduate project for m.s. in cyber operations and resilience, Boise State University, December 2022.

[5] European Marine Observation and Data Network. Emodnet geoviewer. https://emodnet.ec.europa.eu/geoviewer/#, 2025.

[6] Lucía Fernández. Cumulative offshore wind power capacity in europe 2012–2024. https://www.statista.com/statistics/271055/cumulative-european-offshore-wind-power-capacity-installations/, 2025.

[7] Financial Times. Cybersecurity and offshore wind: Industry insights. https://www.ft.com/content/2e3ee55b-f956-415d-9ad0-89d85a6bb260, 2025.

[8] Sarah G. Freeman, Matthew A. Kress-Weitenhagen, Jake P. Gentle, Megan J. Culler, Megan M. Egan, and Remy V. Stolworthy. Attack surface of wind energy technologies in the united states. Technical Report INL/RPT-24-76133, Idaho National Laboratory (INL), January 2024.

[9] Ground Control. Wireless connectivity for offshore wind farms. https://www.groundcontrol.com/blog/wireless-connectivity-for-offshore-wind-farms/, 2023.

[10] Amirashkan Haghshenas, Agus Hasan, Ottar Osen, and Egil Tennfjord Mikalsen. Predictive digital twin for offshore wind farms. *Energy Informatics*, 6(1):1, 2023.

[11] Henry Hui and Kieran McLaughlin. Investigating current plc security issues regarding siemens s7 communications and tia portal. In *Proceedings of the 5th International Symposium for ICS & SCADA Cyber Security Research (ICS-CSR)*. BCS, The Chartered Institute for IT, 2018.

[12] Patrick Katuruza. It-ot convergence: Managing the cybersecurity risks. https://gca.isa.org/blog/it-ot-convergence-managing-the-cybersecurity-risks, 2025.

[13] Anna Knack, Yvonne Kam Hwei Syn, and Kimberly Tam. Enhancing the cyber resilience of offshore wind. Research report, The Alan Turing Institute, Centre for Emerging Technology and Security (CETaS), June 2024.

[14] Obafemi O. Olatunji, Paul A. Adedeji, Nkosinathi Madushele, and Tien-Chien Jen. Overview of digital twin technology in wind turbine fault diagnosis and condition monitoring. In *2021 IEEE 12th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*, pages 201–207, 2021.

[15] Evgeni Sabev, Roumen Trifonov, Georgy Tsochev, Galya Pavlova, and Kamelia Raynova. Analysis of practical cyberattack scenarios for wind farm scada systems. In *2021 International Conference Automatics and Informatics (ICAI)*, pages 420–424, Varna, Bulgaria, September 2021. IEEE.

[16] SecurityScorecard. Ftp security risks, vulnerabilities & best practices guide, August 2025. Accessed: 2025-08-26.

[17] Siemens Energy. Cybersecurity for Wind Offshore. White paper, Siemens Energy, 2024.

[18] Florian Stadtmann, Adil Rasheed, Trond Kvamsdal, Kjetil André Johannessen, Omer San, Konstanze Kölle, John Olav Tande, Idar Barstad, Alexis Benhamoud, Thomas Brathaug, Tore Christiansen, Anouk-Letizia Firle, Alexander Fjeldly, Lars Frøyd, Alexander Gleim, Alexander Høiberget, Catherine Meissner, Guttorm Nygård, Jørgen Olsen, Håvard Paulshus, Tore Rasmussen, Elling Rishoff, Francesco Scibilia, and John Olav Skogås. Digital twins in wind energy: Emerging technologies and industry-informed future directions. *IEEE Access*, 11:110762–110787, oct 2023.

[19] ChatGPT AI Staff. Siemens s7-1500 vulnerabilities in 2025: Risks, impacts, and critical security measures. https://windowsforum.com/threads/siemens-s7-1500-vulnerabilities-in-2025-risks-impacts-and-critical-security-measures.370080/, June 2025. Windows Forum.

[20] Wall Street Journal Staff. European wind-energy sector hit in wave of hacks. *The Wall Street Journal*, 2022.

A MODAT Magnity Service Meta-data

Table 26: MODAT Magnify service schema with wrapped field names and explanations.

Field	**Type**	**Explanation**
ip	string	IP address of the scanned host.
service	object	Container for discovered service details.
transport	string (enum)	Transport protocol (e.g., TCP, UDP).
port	integer	Port number where the service is running.
banner	string | null	Raw service banner (if available).
protocol	string (enum)	Application protocol (e.g., http, ssh, tls).
HTTP		
http	object | null	HTTP details block; null if not applicable.
http.title	string | null	Page title from HTTP response.
http.status_code	integer | null	HTTP status code (e.g., 200, 404).
http.headers	string | null	Raw HTTP headers (deprecated).
http.headers_sha256	string | null	SHA-256 of HTTP headers.
http.body	string | null	HTTP response body (if captured).
http.body_mmh3	integer | null	MurmurHash3 hash of response body.
http.body_sha1	string | null	SHA-1 of response body.
http.body_sha256	string | null	SHA-256 of response body.
http.favicon	object | null	Favicon metadata (if present).
http.css_mmh3	array<integer> | null	(Deprecated) mmh3 hashes of linked CSS.
http.js_mmh3	array<integer> | null	(Deprecated) mmh3 hashes of linked JS.
SSH		
ssh	object | null	SSH details block; null if not applicable.
ssh.hassh	string | null	HASSH fingerprint of the SSH handshake.
TLS		
tls	object | null	TLS/SSL handshake and certificate metadata.
tls.is_self_signed	boolean | null	Whether the certificate is self-signed.
tls.valid_from	string | null	Certificate validity start (ISO date-time).
tls.expires_at	string | null	Certificate expiry (ISO date-time).
tls.supported_versions	array<string>	Supported TLS protocol versions.
tls.fingerprint_sha256	string | null	SHA-256 fingerprint of the leaf cert.
tls.fingerprint_sha1	string | null	SHA-1 fingerprint of the leaf cert.
tls.jarm	string | null	JARM TLS fingerprint.
tls.serial_number	string | null	Certificate serial number.
tls.issuer	object | null	Issuer DN fields (e.g., CN, O, C).
tls.subject	object | null	Subject DN/SANs.
tls.raw	string | null	Raw certificate data (PEM/DER).
tls.is_valid	boolean (read-only)	Whether chain/validity checks pass.
tls.is_trusted	boolean (read-only)	Trusted by the scanner's CA store.

Continued on next page

Field	**Type**	**Explanation**
`scanned_at`	string (date-time)	Timestamp when the service observation was recorded.
Fingerprints & Tech		
`fingerprints`	object | null	Parent container for fingerprint results.
`fingerprints.os`	object | null	OS fingerprint details (product, version, icon).
`fingerprints.service`	object | null	Service fingerprint details (product, version, icon).
`technologies`	array<object>	Detected technologies (name, version, icon).
`tags`	array<string>	Service tags (e.g., `ics`, `vpn`); empty by default.
CVEs		
`cves`	array<object>	Vulnerabilities linked to detected product/version.
`cves.id`	string	CVE identifier (e.g., CVE-2024-12345).
`cves.cvss`	number | null	CVSS severity score if available.
`cves.is_kev`	boolean	In CISA Known Exploited Vulnerabilities (KEV) catalog.
DNS / Network / Geo		
`fqdns`	array<string> (unique)	Fully Qualified Domain Names for the host.
`asn`	object | null	ASN (Autonomous System Number) details.
`asn.number`	integer | null	ASN number.
`asn.org`	string | null	Organization owning the ASN.
`geo`	object | null	Geolocation information for the IP.
`geo.city_name`	string | null	City name.
`geo.country_name`	string | null	Country name.
`geo.country_iso_code`	string | null	ISO country code.
`is_anycast`	boolean (default false)	Whether the IP is an anycast address.

B Role Classification Criterion

B.1 Susbtation Criteria

Condition	**Criteria (match any)**	**Information (fields)**	**Justification**
S1 IEC 61850 stack?	IEC 61850, MMS, GOOSE, SV	Keywords; Tech	Canonical substation automation stack.
S2 Telecontrol to grid?	DNP3, IEC 60870-5-104 / IEC-104 / T104	Keywords; Tech	Collector/transmission control via DNP3/IEC-104 to RTUs/relays.
S3 IED/relay/bay terms?	IED, protection relay, RTU, bay controller; vendors: SIPROTEC, Relion, SEL, MiCOM, Sepam, Multilin	Keywords; Tech	Substations interface with IEDs/relays/RTUs; vendor names add precision.
S4 Grid-coupling vocab?	PoC, PCC, point of common coupling, grid connection	Keywords	PoC/PCC are substation/grid coupling terms.
S5 OPC-UA in substation context?	opc ua/opc-ua/opcua + (SCADA | IED | OSS | substation)	Keywords; Tech	OPC-UA interconnects IEDs/SCADA/control center.
S6 TSO/DSO orgs?	Org includes e.g., TenneT, 50Hertz, National Grid	Organizations	Grid-operator names near OSS are plausible; context-dependent.
S7 OT/SCADA tags?	SCADA, ICS, OT, Power Plant	Tags	Dataset OT tags support substation role.
S8 Substation FQDN hint?	Regex: `\b(oss|onss|substation)\b`	Hostnames	Naming often encodes role; supportive and reviewable.

Table 27: Substation role—conditions, criteria, fields, and justification.

B.2 Control Center Criteria

Condition	Criteria (match any)	Information (fields)	Justification
C1 DMZ/VPN posture?	DMZ, VPN, two-factor/MFA, jump host, bastion	Keywords	Control centers sit behind DMZ/firewalls and support VPN.
C2 CC nouns?	control center, operations center, dispatch, NOC, SCADA center	Keywords	CC aggregates turbine/substation data and supervises ops.
C3 OPC(-UA) without fieldbus?	OPC/OPC-UA present *and no* Modbus, DNP3, S7, IEC-104, PROFIBUS, CAN, . . .	Keywords; Tech	Fieldbuses typical at tower/OSS; OPC-only suggests CC layer.
C4 Enterprise platforms?	Active Directory, ISE, DNA Center, WLC, SIEM, FMC/Firepower	Keywords; Tech	Enterprise tooling concentrates at CC/enterprise, not at field.
C5 Remote-access nouns?	remote access, MFA, gateway, bastion, jump host	Keywords	Reinforces C1; consistent with CC exposure patterns.
C6 Operator orgs?	Org includes Orsted, Iberdrola, Vattenfall (examples)	Organizations	Operator orgs near CC assets are plausible; use with context.
C7 Tags fit CC?	SCADA, ICS, Remote Access, Cloud	Tags	Supervisory/remote-access posture tags.

Table 28: Control Center role—conditions, criteria, fields, and justification.

B.3 Not Part OWF Criteria

Condition	Criteria (match any)	Information (fields)	Justification
N1 (hard) Non-wind tags?	Solar Panel, Charging Station, Healthcare, TV Streaming, BitTorrent, 3D Printer, Parking System, Printer, Home Automation, Smart Garden	Tags	Domain mismatch → immediate exclusion.
N2 (soft) Shared-hosting FQDNs?	Hostnames like `^cpanel.`, `^cpcontacts.`, `^webdisk.`, `^webmail.`, `^mail.`; contains `plesk/webhost/wixdns`	Hostnames: `fqdns`	Shared hosting/web panels rarely map to OWF OT; reviewable heuristic.
N3 (soft) Generic IT/CMS/webmail UIs?	`wordpress`, `roundcube`, `webmin`, `drupal`, `joomla`, `cpanel`, `plesk`	Keywords; Tech	Commodity IT admin UIs → likely not OWF OT; reviewable heuristic.

Table 29: Not Part of OWF—exclusion conditions, criteria, fields, and justification.