12-10-2021

Leveraging choice modeling technique for enhancing the cyber resilience of the smart grid

Kesava Karishma Devi Dadi
Mississippi State University, karishmadadi@gmail.com

Leveraging choice modeling technique for enhancing the cyber resilience of the smart grid

By

Kesava Karishma Devi Dadi

Approved by:

Raed Jaradat (Major Professor)
Haifeng Wang
Niamat IbneHossain
Linkan Bian (Graduate Coordinator)
Jason M. Keith (Dean, Bagley College of Engineering)

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Industrial and Systems Engineering
in the Department of Industrial and Systems Engineering

Mississippi State, Mississippi

December 2021

Name: Kesava Karishma Devi Dadi

Date of Degree: December 10, 2021

Institution: Mississippi State University

Major Field: Industrial and Systems Engineering

Major Professor: Raed Jaradat

Title of Study: Leveraging choice modeling technique for enhancing the cyber resilience of the smart grid

Pages in Study 36

Candidate for Degree of Master of Science

This research focuses on the cyber-attack of the smart grid and its retrieval to a normal state by estimating the smart grid's resilience. This study developed a theoretical model to estimate the resilience of the smart grid using choice modeling. A utility function is formulated based on various factors and subfactors of resilience to estimate the resilience of the smart grid. Choice modeling is applied to estimate the model parameters in various fields such as marketing, energy, transportation, and health and to predict the outcome.

DEDICATION

I would like to dedicate this work to my parents Dadi Nageswara Rao and Baby, and my brothers Venkatesa Raju and Dinesh Kumar for their continuous support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The smart grid supplies electricity to stations in the centralized era and transmits them to customers through transmission and transmission systems. The network is operated, controlled, and monitored using information and communications technology (ICT). These innovations allow energetic companies to continuously monitor their control requirements and gain the capacity and reliability of control shipments at a lower cost. The smart grid framework provides the most qualified electricity organization operations based on the information obtained by consumers through improved two-way communication between buyers and electricity control companies. Security is a top vital concern of smart grid systems because of the risks and burdens those residents and businesses might encounter of an attack on the power grid. Three fundamental security goals must be joined within the smart grid framework:

1) the ability to provide uninterruptible power on demand of the user;

2) the integrity of the information communicated;

3) the confidentiality of the information Confidential data on the user's network (Aloul et al., 2012).

Within the smart grid environment, the characteristics of control systems can be summarized as extendible, open, and available. Since these characteristics make such systems more helpless, in any case, it is critical to get the nature of cyber dangers in arrange to guarantee that they are secure

(김발호.,2016). A cyberattack is the loss or destruction of a targeted executive by modifying or redirecting stored information using various attack procedures on the Internet. In ICT, cyber risks are basically classified into system, network, and web level risks. As a prerequisite for a cyber-attack, specific data on the target frame must be collected during the so-called information retrieval phase. System-level, network-level, Web-level, and information seeking stage (김발호,2016). The Smart Grid addresses a phenomenal chance to move the energy business into another time of unwavering quality, accessibility, and effectiveness that will add to our financial and natural wellbeing (U.S department of energy, Smart grid).

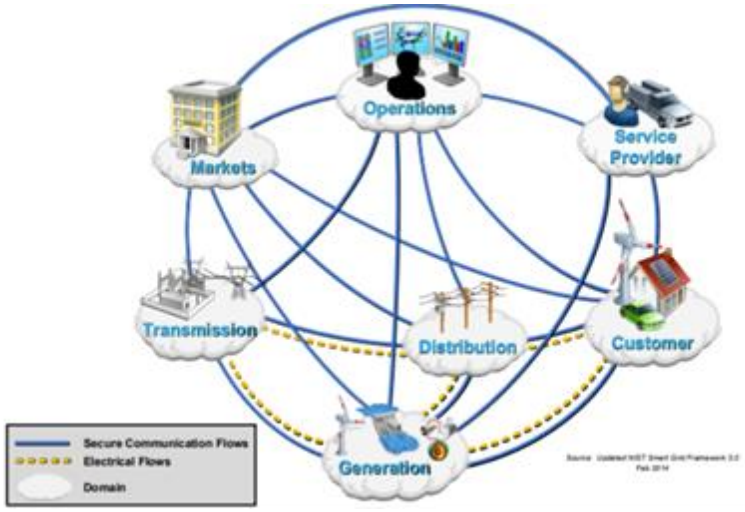

Figure 1 NIST Smart Grid Conceptual Model-Interaction of roles in different smart grid domains through secure communication (FitzPatrick,2011).

Resilience of the smart grid

Resilience is the potential of a system to quickly and efficiently recover from an adverse condition impacting the system's performance to decrease and resume regular operation. Extreme catastrophes such as hurricanes, tornadoes, and floods, as well as large-scale coordinated cyber-

attacks and ways to recover from these occurrences, are often addressed while examining grid resilience. Designing microgrids that can run independently of each other and thus provide a way for quick recovery from an extreme occurrence outage is a frequent strategy used to improve smart grids' resilience (Das et al,2020). The ability of a system to tolerate interruptions while maintaining its form and composition is also known as resilience (Fiksel,2003).

Choice Modeling helps in the analysis of questions such as "What, why, and how the behavior of the factors affects the circumstance or encountered problem," as well as providing a consequence with predicted outputs using various modeling methodologies across numerous disciplines across time. We can identify major factors or variables in consumption or repetitive behavior by studying the outcomes. In sectors such as marketing, transportation, energy, and health, choice modeling is used to estimate model parameters (S Hess et al,2020). For example, Individuals make numerous decisions in transportation, such as whether to travel or not, which method of travel to use, and which route or transit line to use. All of these judgments entail picking one option from a list of several: a discrete choice (Molly et al,2019).

In business, resilience is defined as an organization's, resource's, or structure's ability to withstand the effects of a system failure, retrieve, and restart operations in order to continue to deliver essential services (Hoffman,2007). This definition of choice modeling has implanted the idea and strength to the idea of implementing choice modeling to the smart grid. As withstanding of an attack is the primary function of smart grid in terms security with increase of cyber-attacks.
Till today, many different techniques have been implemented to calculate the resilience of the smart grid. This is the first research to propose a choice modeling technique (estimation) to

increase the resilience of smart grid by calculating the resilience as the ratio of recoverability to vulnerability (Ibne Hossian et al,2020). In this research a conceptual framework is designed to estimate the resilience with effect of various factors during attack using multi-nominal logit and Nested logit and hessian matrix to calculate the probabilities of the likelihood function.

Section 2 of this research talks about the literature review of the resilience capacity, section 3 defines choice modeling, section 4 describes the methodology of the framework and section 5 concludes with summary and limitations of the research.

CHAPTER II

LITERATURE REVIEW

Vulnerability and restoration are parent nodes of resilience (Ibne Hossian et al,2020). According to Ibne Hossain et al(2020), resilience is calculated as a ratio of recoverability(restoration) to vulnerability. Equation 1 below represents the resilience calculation:

$$\boldsymbol{Resilience} = \frac{\boldsymbol{Recoverability(restoration)}}{\boldsymbol{Vulnerability}} \quad (1)$$

The mechanisms for recovering a geographic area/organization from any impact or exterior disturbances due to interruption are known as resilience capacities. The absorptive capacity, adaptive capacity, and restorative capacity of the system can be used to describe smart grid restoration (recoverability). The three components of the software, access, and network are now used to solve smart grid vulnerability (Ibne Hossian et al,2020). The parameters of the proposed technique are specified in this research from the aspect of cyber resilience. In order to effectively analyze the cyber resilience of the smart grid system, it is vital to first identify the most important parameters. The following approach is used to determine most of the relevant factors associated with vulnerability and recoverability. Within the framework of the smart grid system, a comprehensive study has been made, examined, and expert opinions were incorporated. The sources of vulnerability, as well as several methods for restoring the smart grid from hackers, are explored in the sub-sections below.

Restoration(recoverability)

Restoration (recoverability) can be represented using a specific type of resilience capacities, comprising absorptive, adaptive, and restorative capacity (Biringer et al, 2013). A system's resilience capacity is an internal characteristic that enhances the ability to absorb, adapt, and recover from any foreign attacks or interruptions (Hossian et al,2019). The following is an overview of various resilience capacities.

Absorptive Capacity

Absorptive capacity is an internal aspect of the syst(em that is mostly regarded to be the first line of protection against the disruption's consequences Biringer et al,2013; Hossian et al,2019). The system's absorptive capacity refers to a series of explicit proactive actions through which a system will automatically cope with the shock's vulnerability or sensitivity with little effort. Advanced metering infrastructure (AMI) and Visualization technology are the two important factors of absorptive capacity of the smart grid resilience (Ibne Hossian et al,2020).

Advanced metering infrastructure (AMI)

AMI enables providers a two-way communications network with meters and also the opportunity to modify service-level parameters for users (Farhangi,2009). AMI, also referred to as "smart metering," is a key element of the smart grid system, which includes smart meters, concentrators, and the Meter Data Management System, which together enables effective encrypted connection, energy consumption measurements, connectivity with outside nodes, data storage, and management, among several other things (Mohammadali et al,2016). According to McLaughlin et al., an attacker's method of tricking the energy grid by affecting AMI systems and conduct security audits on commodity devices to confirm the effectiveness of these attacks. They led to the

realization that not only is theft viable in AMI systems, but that modern AMI devices present a variety of new channels to do so (S. McLaughlin et al,2009). Varodayan and Gao (2010) suggested a secure mechanism for distribution companies to reflect the energy readings they receive from smart meters back to consumers so that users may check the smart meter's integrity. This would restrict the attacker from manipulating the smart meter reading and guarantee the reading accuracy. The integrated advanced metering management system was first implemented in 2004 to maximize energy efficiency in Italy's capital. The system features high accuracy trans meters and smart grid applications, including network operation control and the capability to monitor low and medium voltage line status automatically (Fang et al,2011). The key elements of AMI are the HAN, smart meter, operational gateway, and meter data management system. AMI collects, evaluates, stores, and transmits measurement data from smart meters to authorized parties. As a result, they may use the data to estimate demand, manage outages, and bill customers. It assists consumers in maximizing energy efficiency by providing real-time electricity prices (Gunduz & Das,2020). Figure 2 shows a conventional communications infrastructure topology for smart grid AMI based on the literature (A. Hamlyn 2008; G.A.Taylor,2006).

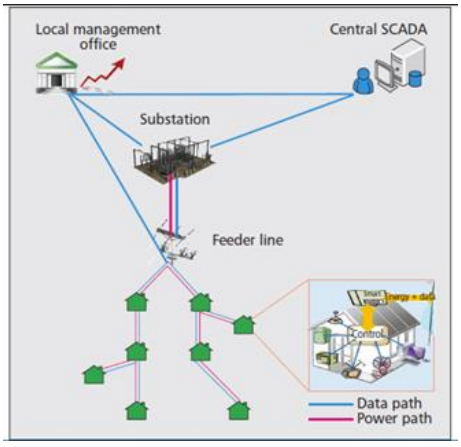

Figure 2 A smart grid AMI communication network architecture

The multitude of meter reading messages to be validated may overload a smart grid system that connects dozens of buildings, each with many units. Furthermore, smart grid AMI communications are vulnerable to traditional cyber assaults like capacity depletion and disguise attacks due to wireless and IP technology deployment. The requirements of AMI security are:

- Authentication of device
- Confidentiality of data
- Message Integrity
- Maintaining confidentiality of some parts
- Avoiding potential cyber-attacks.

Digital signature and authenticating each message can improve communication security; approaches based on traditional cryptographic operations are neither effective nor extensible in a smart grid system due to traffic volume and resource limits. To perform meter reading data collecting and management message dissemination securely and efficiently, we need minimal but efficient and reliable techniques designed specifically for smart grid AMI communications (Yan et al,2013).

Visualization Technology

Experts need to integrate cutting-edge technologies into the structure of smart grid technology to outsmart the actions of cyber attackers (Venugopalan & Rai,2015). At the utility level, grid visualization tools can be used for real-time load monitoring and load growth planning. When the demand response program for users grows, it may be difficult to interpret and visualize data. The VERDE platform, embedded in Google Earth, enables wide-area grid monitoring by combining real-time sensor data, weather forecasts, and grid modeling with geographic data to allow the smart

grid system cyber protected. It also can provide real-time information regarding outages and power quality to improve system reliability (Ibne Hossian et al,2020). In terms of power system operation, forecast, fault detection, dispatch, and service, visualization provides an excellent interface to the power system operator (Chen& Chen,2021). Low dimensional techniques, multivariate high dimensional and Geographical information system (GIS) techniques are three main categories of techniques used to visualize the smart grid data distribution.

According to M. Stefan et al., Time-critical (grid monitoring) and non-time-critical (grid history) are two features of visualization. In grid monitoring, the system's real-time visualization capabilities are heavily dependent on the frequency at which data is gathered and saved in the database, as well as the occurrences during real-time data processing. To obtain the shortest processing delays in this situation, one must determine what data must be analyzed by each visualization aspect. For the tracking system, the latency in data processing is crucial. It must be minimized while calculating the amount of data necessary to make competent conclusions, the cost of data monitoring and generated events, and the time it takes to retrieve the important information Ex: Voltage drop/rise. Non-time-critical characteristics are associated with electrical grid planning, which is used to develop a model for future state estimations. As a result, what parameters to measure, which data to pass onto the visualization system, and how to analyze the data should be determined so the database visualization system can be optimized Ex: Power Balancing (M Stefan et al,2017). Various attributes of nodes in the system can be visualized to yield a better understanding of the nature of a cyberattack (Matuszak et al,2013). Table 1 represents various tools used in visualization technology.

Table 1 Various tools used in Visualization Technology

Author	Tools	Summary/Goal
Foreman et al (2015)	Immersive Visualization tool	Provides users with a simulation that can support both existing and proposed power grid technologies while engaging them in a realistic visualization environment
Nga, D. V et al (2012)	Single line diagram, 2D chart and 3D surface with contour	Single line diagrams show how substations are connected as well as other important parameters such as line status and AMI energized condition. 2D Chart: A bar chart is used to depict dynamic data. The number of data points is represented on the x-axis, while the total energy consumption of all substations is represented on the y-axis. 3D Chart: Grid operators utilize 3-D displays to visualize power usage or load in 1-hour intervals over the course of the month.

Table 1(continued)

Author	Tools	Summary/Goal
Chen, X et al (2021)	Virtual reality (VR) and Augmented reality (AR)	Virtual reality (VR) is a significant technology for improving user interaction with virtual scenes in future data visualization. Augmented Reality (AR) technology is to overlaps virtual information on physical objects to create a new virtual reality world.
Nga, D. V et al (2012)	Parallel coordinate, scatter diagram and Andrew curve	Instead of employing orthogonal axes as in a traditional Cartesian graph, the coordinate axes in parallel coordinates are all set down horizontally. In the plot, each observation is represented by a set of connected line segments. The scatter diagram shows all the bivariate scatter plots between the project's five variables, as well as a uni-variate distribution for each variable. The Scatter plot displays only bi-variate relationships. Each sample is represented as a smooth function over the interval [0, 1] in the Andrew curve plot.

Table 1 (continued)

Author	Tools	Summary/Goal
Matuszak et al (2013)	CyberSAVe visualization framework	CyberSAVe enables a system operator to determine not just the existence of hardware or software failure but also the reason behind the failure and, possibly, prevention methods.
Khan, M.I.U et al (2016)	Spatial Analysis	The additional and new information from the GIS data is extracted via spatial analysis. Different geographic tools were employed in GIS for feature statistics, as well as a buffer, intersect, union, and other geoprocessing tools in spatial analysis for smart grid systems.

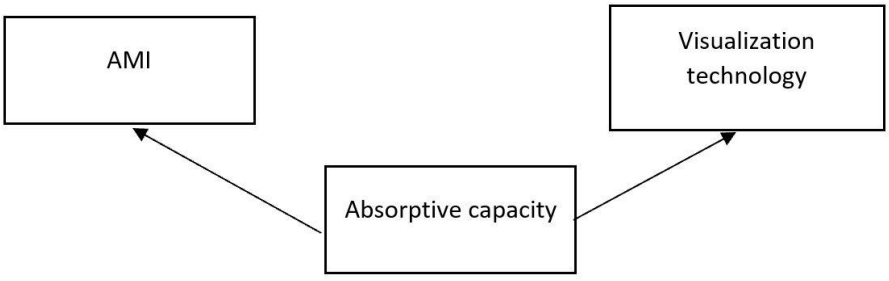

Figure 3 Two factors of absorptive capacity considered to construct the utility function

Adaptive Capacity

The potential of a system to self-organize and provide timely responses to deal with external disturbances without any restorative activity is described as adaptive capacity, which is regarded as the midline of defense (Ibne Hossian et al,2020). Enhancing the system's adaptive capacity, either by ensuring that the system is designed with enough redundancy to continue to function or by boosting the system's ability and pace to modify and adapt to change as they arise. '*Strategy, operations, management systems, governance structure, and decision-support capabilities'* to tolerate interruptions and perturbations is defined as adaptive capacity (Dalziell&McManus,2004). Grid partitioning and delay adaptive control strategy are the two important factors of adaptive capacity (Ibne Hossian et al,2020).

Grid partitioning

Grid partitioning enables system operators to alter voltage regulation within each microgrid and reduce signal security concerns and compounding errors in the vastly complicated network during any disturbance, reducing the impact of cyber disruptions (Ibne Hossian et al,2020).

Delay adaptive control

A delay-adaptive control technique can be implemented to develop a resilient smart power grid, reducing the network latency of the system to a delay-free process (Ibne Hossian et al,2020).

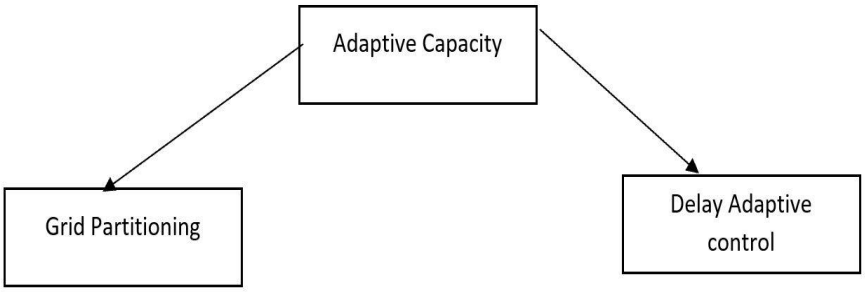

Figure 4 Two factors of adaptive capacity considered to construct the utility function

Restorative capacity

By comparison to other recovery methods, cyber control recovery is projected to be faster. System restoration may take longer than expected if the malicious virus infects the entire SG system of systems. It will take a sophisticated team of professionals who can work on an advanced decision support platform to ensure that every infected machine is rapidly restored to its original state. The restorative capacity comprises of two main factors namely restoration of control and restoration of self-healing in calculating the resilience of the smart grid (Ibne Hossian et al,2020).

Restoration of control

Smart grid control architecture is defined as restoring control for fault diagnosis and power retrieval

Restoration of self-healing

A malfunction may occur within a power underlying system switch breaker, which can be fixed using a restorative self-healing process like an artificial immune response as an optimization technique (Ibne Hossian et al,2020).

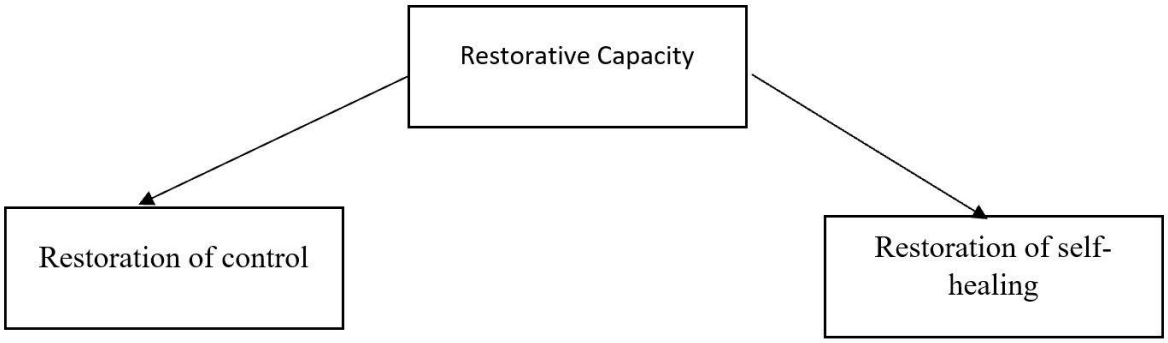

Figure 5 Two factors of restorative capacity considered to construct the utility function

Vulnerability

The performance of the smart grid is impacted by three domain vulnerabilities, namely access domain, network domain, and software domain vulnerability. Each domain vulnerability is being further impacted by more than one factors (Ibne Hossian et al,2020).

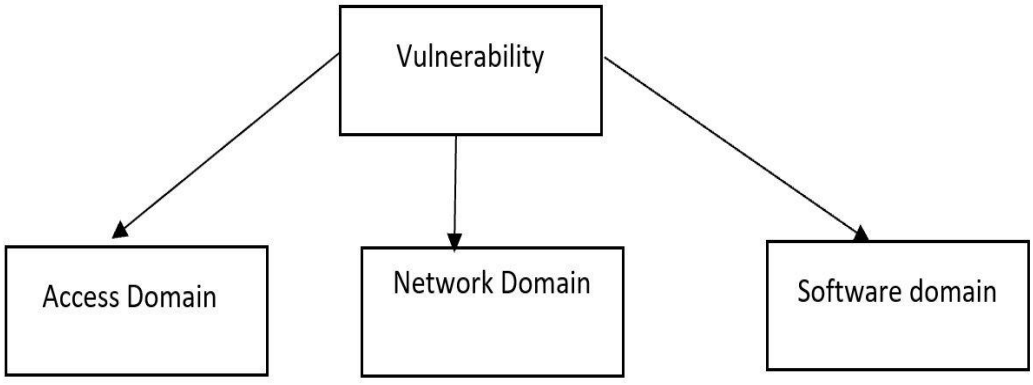

Figure 6 Three factors of vulnerability impacting the smart grid

Access domain Vulnerability

Many common cyber vulnerabilities of the access control domain include the failure to remove duties through allocated access privileges, the failure to restrict system administration for unsuccessful attempts, and the inability to end remote access sessions after a predetermined duration. Weak user and Unauthorized protocols compromise the security of the smart grid leading to access domain vulnerability (Ibne Hossian et al,2020).

Network domain Vulnerability

The architectural development and execution techniques are linked to vulnerabilities in the network domain. From the network domain, a correct proposed system can monitor a process and regulate the delivery of computation for a business function. Network configuration and network audit and monitor can be considered as to-priority in network domain vulnerability (Ibne Hossian et al,2020).

Software domain vulnerability

There are several underlying factors causes software domain vulnerability. Weak code and improper data validation are few important factors affecting the software domain (Ibne Hossian et al,2020).

CHAPTER III

CHOICE MODELING

The main objective of Choice Modeling is to comprehend the behavioral hypothesis in order to predict, enhance, and implement quantifiable ways to construct a model. To construct a model in choice modeling, we identify the dependent variables first, then determine the relationship between variables, predict the likelihood, and estimate the parameters using the data in the database. Alternative specific constants are calculated using a mathematical equation known as the Utility function for available variables. Choice Modeling has been implemented using a variety of software in recent years such as Python Biogeme (Bierlaire,2018), apollo package in R etc (Hess & Palma,2019).

The theory-driven models have dominated choice modeling filed. Choice models that are driven by behavioral theories, such as Utility Theory (UT) (McFadden 2001), Regret Theory (RT) (Loomes and Sugden 1982), and Prospect Theory (PT) (Kahneman and Tversky 1979; Tversky and Kahneman 1992), are examples of theory-driven discrete choice models. Logit, probit, mixed logit, and latent class models are commonly used to estimate complex models. The goal of choice modeling researchers is to build models that generate interpretable, repeatable, scalable, versatile (including the required complexity), and robust predictions and findings (in out-of-sample, externally valid). Choice modeling concepts include random variable, probability distribution,

error term, confidence and prediction intervals, estimator consistency and asymptotic features, central limit theorem, and latent variable (Van Cranenburgh et al,2021).

The actual data generation process – which in the case of choice modeling is the decision-making process– is a stochastic process that may be discovered by evaluating alternative statistical models, according to a choice modeler. Competing models for a choice modeler come from a variety of decision theories, such as Utility Theory, Prospect Theory, Regret Theory, and functional forms, such as different error term distributions. The analyst considers the optimal model to be the actual representation of the data from which behavioral inferences can be drawn after recognizing it. That is, the model's parameters are interpreted in terms of the meaning offered by behavioral theory (assuming the model is valid). A logit kernel, a membership function, a mixing kernel, a random number generator, and other well-known building blocks are commonly used in choice models. These components are included in several popular software packages, including Biogeme, Pylogit, and Apollo. (Van Cranenburgh et al,2021).

For decades, discrete choice modeling is used to model and predict choices as a mathematical tool. Although complicated model structures were developed theoretically in the early phases, it wasn't until recent increases in computing power that this technique became widely used and enabled future advancements (Calastri,2020). Multinominal Logit (MNL) and Nested Logit (NL) are the widely used discrete choice models techniques to estimate the parameters in various environments. The theory of behavior predicts how an individual behaves but not how they should and the aptness of the model to apply in various situations in real-time functioning. To construct a behavioral model, we have to take into a picture of decision-maker choice; in general, a decision-maker can

be an individual or a group of persons or firm with a variety of tastes, behavior, and situations having different socio-economic characteristics such as age, income, level of education, employment to understand their choices from available alternatives by identifying the variables to describe each alternative. Decision maker of the model, denoted by *n*. Need to list the choice set, either a continuous or a discrete set, define and characterize the alternatives using different variables. Decision-maker decision is represented in the form of a mathematical function known as 'utility function' involving the alternatives of choice set with a value, we select the alternative with the highest utility value to make decisions, and decision-maker is assumed to be optimizer and finds an optimized solution. The optimal solution differs with the change of different parameters. To find optimal quantity as a function of parameters, we use the demand function. In specific scenarios, the utility function depends on both continuous goods quantity and discrete choice function made. The Utility function is represented as U_n and Choice set is represented by C_n. In theoretical development, the behavior of an individual is assumed to be deterministic, with assumed properties of the consistent preferences, which is not valid in real scenarios as preferences are inconsistent. To understand the observations of inconsistent preferences, the probabilistic choice theory is used. In probabilistic choice theory, the probability of different outcomes is determined by a specific decision-maker and situation. In developing the model, there are two assumptions made firstly, humans are deterministic and choose the highest utility alternative. Second, the analyst is not sure of utility as he doesn't have a clear view of the factors affecting human behavior and treats it as a random variable. This is known as a random utility approach, combining the approach or idea for building, and operating the model (Bierlaire,2003).

Probabilistic model is expressed as

$$p(i|C_n) = p_\gamma\left(U_{in} \geq U_{jn}, \forall \epsilon C_n\right) \quad (1)$$

Where,

$$U_{in} = v_{in} + \varepsilon_{in} \quad (2)$$

U_{in} → Utility; V_{in} → Deterministic part of utility observed by analyst; ε_{in} → Error term;

Multi-nominal Logit

Multinomial logit models have been used to explain how decision makers compare and assess alternatives since the 1970's (McFadden,1973). The IIA property of MNL states that the ratio of the probabilities of selecting any two options is independent of the availability of the third alternative (Yun&Sun,2012). One of the most appropriate techniques to study preferences for mode choice is the random utility maximization (RUM) framework, which is a discrete choice model. Individual n chooses choice i when the related utility is the highest when compared to the other options, it is predicted (Ashkrof et al,2019).

The utility function of the multinomial logit consists of two parts namely deterministic and error term (S Hess et al,2020).

$$U_{in} = v_{in} + \varepsilon_{in} \quad (4)$$

Where, U_{in} → Utility; V_{in} → Deterministic part of utility observed by analyst; ε_{in} → Error term. (Bai et al.,2017).

In MNL model, the probability for alternative i in choice task t for person n is given by:

$$P_{n,i,t}(\beta) = \frac{z_{avail,i,n,t} * e^{v_{i,n,t}}}{\Sigma_{j=1}^{\bar{J}} z_{avail,j,n,t} * e^{v_{j,n,t}}}. \quad (3)$$

Where β is a vector combining all the parameters and $V_{j,n,t}$ refers to the deterministic part of utility function and $e^{v_{j,n,t}}$ refers to the error term. No limitations are being exposed on the flexibility.

The likelihood function of the model applied is

$$L_n(\beta) = \prod_{t=1}^{Tn} P_{j_n^*,t} \quad (6)$$

Where Tn is the number of separate choices set situation

Nested Logit

The nested logit model is based on the notion that some choices can be combined into many groups (called nests). In the same nest, the error terms may show some correlation, although error terms from other nests are still uncorrelated (Yun&Sun,2012). Nested logit model segregates the utility function by the nesting specifications at different levels within the nested probabilities. The root of the nested parameter is normalized to 1 (S Hess et al,2020).

The probability of choosing alternative i in situation t by person n is given by:

$$P_{i,n,t} = P_{m,n,t} P_{(o_m|m),n,t} P_{(i|o_m),n,t} \quad (7)$$

Where,

$$P_{(i|o_m),n,t} = \frac{e^{\left(\frac{v_{i,n,t}}{\lambda_0 m}\right)}}{\Sigma_{j\in 0} m^{e\left(\frac{v_{j,n1^t}}{\lambda_0 m}\right)}} \quad (8)$$

$$P_{(o_m|m),n,t} = \frac{e^{\left(\frac{\lambda_0 m}{\lambda m} I_{0,n,t}\right)}}{\Sigma_{L_m=1}^{M} e^{\left(\frac{\lambda_l}{\lambda m} \mathrm{Im},n,t\right)}} \quad (9)$$

$$P_{m,n,t} = \frac{e^{\left(\frac{\lambda m}{\lambda r} Im,n,t\right)}}{\Sigma_{m=1}^{M} e\left(\frac{\lambda_L}{\lambda_\gamma} I_l, n, t\right)} \quad (10)$$

Table 2 Represents various applications of MNL and NL models

Author	Application	Model
Hagenauer et al (2017)	Travel mode	MNL
Alwosheel et al (2019)	Travel mode	MNL
Mohammadian & Miller (2002)	Vehicle Type	NL
Newman et al (2020)	Airline Itinerary	MNL, NL
Yao et al (2020)	Route	MNL

CHAPTER IV

METHODOLOGY

The approach for the proposed methodology was developed after a thorough examination of the literature. The search for available literature was guided by keyword phrases (i.e., smart grid, resilience capacities, adapt, recover, vulnerability, etc.) referring to the factors and subfactors impacting smart grid resilience in the Scopus and Web of Science databases. The database search comprised peer-reviewed publications, websites, conferences, and book chapters in order to understand all issues linked to the resilience of the smart grid and cyber-attacks on the smart grid, as well as choice modeling techniques on diverse applications. To increase the resilience of the smart grid, 12 attributes are chosen upon extensive literature review.

In this research, we have designed the conceptual model for estimating the resilience of the smart grid using the apollo package in R (S Hess et al,2020). We have followed the following steps to formulate the utility function for estimating resilience.

Step 1: Identified the factors affecting the resilience of the smart grid with an extensive literature review. With the literature review, we observed Adaptive, Absorptive, and restorative capacity plays a key role in calculating the resilience of the smart grid along with vulnerability factors.

Step 2: Quantified factors and sub-factors affecting the resilience of the smart grid. MS Excel is used to quantify the factors and sub-factors based on the Ibne et al., (2020) paper modeling techniques, depict the result as following for restorative and vulnerability

Table 3 Represents quantified data representation of factors and sub-factors

Factors	Description	MS Excel
Absorptive Capacity		
AMI	Cyber-attack prevented – True; otherwise - False	1-True; 0- False
Visualization Technique	Failure prevented -True; otherwise - False	1-True; 0- False
Adaptive Capacity		
Grid Partitioning	Failure prevented -True; otherwise - False	1-True; 0- False
Delay-adaptive control strategy	Varies between 80 and 120ms with mean delay 100ms	Mean ≈ 100ms
Restorative Capacity		
Restoration of Control	Retrieved to the fullest – True; Otherwise- False	1-True; 0- False
Restorative Self-Healing	Self-healed- True; Otherwise - False	1-True; 0- False
Software domain vulnerability		
Weak code	Weak code – True, Otherwise - False	1-True; 0- False
Improper data validation	Improper validation- True, Otherwise - False	1-True; 0- False
Access domain vulnerability		
Weak User	Weak User- True, Otherwise - False	1-True; 0- False
Unauthorized protocols	Unauthorized protocol- True, Otherwise - False	1-True; 0- False
Network domain vulnerability		

Table 3 (continued)

Network Configuration	Safe configuration- True, Otherwise - False	1-True; 0- False
Network audit and monitoring	Weak Network- True, Otherwise - False	1-True; 0- False

Step 3: Define the model parameters, definition, and Estimation.

In this model, we consider utility function U according to equation 4, for each factor, denoted by i ∈ ***I*** in situation denoted by t ∈ ***T*** for each circumstance/ attack n ∈ N are given by

$$U_{i,n,t} = \delta_i + \beta_1^* x_1 + \beta_2^* x_2 + \varepsilon_{j,n,t} \quad (11)$$

Where, δ_i represents the alternative constant for each factor for both vulnerability and restorative factors, x_1 and x_2 denotes sub-factors of each factor. β_1 and β_2 denotes the combining vector of each sub-factor. $\varepsilon_{j,n,t}$ denotes the error term of the utility function. The utility function has two terms deterministic and error term of each factor, deterministic term consists of sub-factors of each factor. Equation 11, can be further simplified for recoverability and vulnerability as follows:

Utility functions to calculate the recoverability capacity of the smart grid are:

1. Utility function of absorptive capacity denoted by $U_{ab,n,t}$:

$$U_{ab,n,t} = \delta_{ab} + \beta_1 * x_{ab1} + \beta_2 * x_{ab2} + \varepsilon_{ab,n,t} \quad (12)$$

Where, δ_{ab} is specific constant of absorptive capacity, x_{ab1} represent quantified value of AMI and x_{ab2} denotes Visualization technology Boolean value and $\varepsilon_{ab,n,t}$ is the error term of absorptive capacity. β_1 and β_2 denotes the Combining vector of each sub-factor. Every utility function comprises of two terms – deterministic and error term.

2. Utility function of adaptive capacity denoted by $U_{ad,n,t}$:

$$U_{ad,n,t} = \delta_{ad} + \beta_1 * x_{ad1} + \beta_2 * x_{ad2} + \varepsilon_{ad,n,t} \quad (13)$$

Where, δ_{ad} is specific constant of absorptive capacity, x_{ad1} represent Grid partitioning quantification value as represented in quantification table in step 2 and x_{ad2} denotes delay-adaptive control strategy with values in range of 80-120ms and mean 100ms generated using MS Excel and $\varepsilon_{ad,n,t}$ is the error term of adaptive capacity. β_1 and β_2 denotes the Combining vector of each sub-factor. Every utility function comprises of two terms – deterministic and error term.

3.Utility function of restorative capacity denoted by $U_{rs,n,t}$:

$$U_{rs,n,t} = \delta_{rs} + \beta_1 * x_{rs1} + \beta_2 * x_{rs2} + \varepsilon_{rs,n,t} \quad (14)$$

Where, δ_{rs} is specific constant of restorative capacity, x_{rs1} represent quantified value of retrieve of control and x_{rs2} denotes restorative self- healing Boolean value and $\varepsilon_{rs,n,t}$ is the error term of restorative capacity. β_1 and β_2 denotes the weights of each sub-factor. Every utility function comprises of two terms – deterministic and error term.

Utility functions to calculate the vulnerability of smart grid are:

1. Utility function of software domain vulnerability $U_{sd,n,t}$:

$$U_{sd,n,t} = \delta_{sd} + \beta_1 * x_{sd1} + \beta_2 * x_{sd2} + \varepsilon_{sd,n,t} \quad (15)$$

Utility function comprises of deterministic and error term of the utility function, where δ_{sd} represent specific constant of software domain vulnerability and x_{sd1} represent weak code quantification value as represented in quantification table in step 2 and x_{sd2} denotes improper data validation Boolean value generated using MS Excel and $\varepsilon_{sd,n,t}$ is the error term. β_1 and β_2 denotes the Combining vector of each sub-factor.

2. Utility function of access domain vulnerability $U_{av,n,t}$:

$$U_{av,n,t} = \delta_{av} + \beta_1 * x_{av1} + \beta_2 * x_{av2} + \varepsilon_{av,n,t} \quad (16)$$

Where, δ_{av} is specific constant of access domain vulnerability, x_{av1} represent quantified value of weak user and x_{av2} denotes unauthorized protocol Boolean value and $\varepsilon_{av,n,t}$ is the error term of access domain vulnerability. β_1 and β_2 denotes the Combining vector of each sub-factor. Every utility function comprises of two terms – deterministic and error term.

3. Utility function of network domain vulnerability:

$$U_{nd,n,t} = \delta_{nd} + \beta_1 * x_{nd1} + \beta_2 * x_{nd2} + \varepsilon_{nd,n,t} \quad (17)$$

Utility function comprises of deterministic and error term of the utility function, where δ_{nd} represent specific constant of software domain vulnerability and x_{nd1} represent safe configuration quantification value as represented in quantification table in step 2 and x_{nd2} denotes weak network Boolean value generated using MS Excel and $\varepsilon_{nd,n,t}$ is the error term. β_1 and β_2 denotes the Combining vector of each sub-factor.

Table 4 Represents the parameters of proposed utility function

Parameters	Description
$U_{ab,n,t}$	Utility function of absorptive capacity at a circumstance n and time t.
δ_{ab}	Specific constant of absorptive capacity
x_{ab1}	AMI quantification value
x_{ab2}	Visualization Boolean value
$\varepsilon_{ab,n,t}$	Error term of absorptive capacity
$U_{ad,n,t}$	Utility function of adaptive capacity at a circumstance n and time t.
δ_{ad}	Specific constant of adaptive capacity
x_{ad1}	Grid Partitioning
x_{ad2}	Delay- adaptive control
$\varepsilon_{ad,n,t}$	Error term of absorptive capacity
$U_{rs,n,t}$	Utility function of restorative capacity at a circumstance n and time t.
δ_{rs}	Specific constant of restorative capacity
x_{rs1}	Restorative of control value
x_{rs2}	Restorative self-healing value
$\varepsilon_{rs,n,t}$	Error term of restorative capacity

Table 4(Continued)

$U_{sd,n,t}$	Utility function of software domain vulnerability at a circumstance n and time t.
δ_{sd}	Specific constant of absorptive capacity
x_{sd1}	Weak code
x_{sd2}	Improper data validation
$\varepsilon_{sd,n,t}$	Error term of absorptive capacity
$U_{av,n,t}$	Utility function of access domain vulnerability at a circumstance n and time t.
δ_{av}	Specific constant of access domain vulnerability
x_{av1}	Weak user
x_{av2}	Unauthorized protocol
$\varepsilon_{av,n,t}$	Error term of access domain vulnerability
$U_{nd,n,t}$	Utility function of network domain vulnerability at a circumstance n and time t.
δ_{nd}	Specific constant of network domain vulnerability
x_{nd1}	Network configuration – safe or unsafe
x_{nd2}	Weak network – Audit and Monitoring of network
$\varepsilon_{nd,n,t}$	Error term of network domain vulnerability
β_1 and β_2	Combining vector of sub-factors

We can consider utility function with two possible cases with user initializing initial value of parameters:

Case 1: Can fix the values of certain parameters over course of time allowing no changes to initial value stored in it.

Case 2: Value of parameter can keep changing in the course of implementation with initialization of the parameter with start value.

Initial values of defined parameters input can also include the outputs of any other or previous model for model that is complex or hybrid.

Multi nominal Logit model

In above utility functions $\varepsilon_{ab,n,t}$, $\varepsilon_{ad,n,t}$, $\varepsilon_{rs,n,t}$, $\varepsilon_{sd,n,t}$, $\varepsilon_{av,n,t}$, $\varepsilon_{nd,n,t}$ denotes error term distributed identical and independent values over different factors effecting the resilience value with type I extreme distribution value.

MNL model is obtained with the probability for each factor i at the given time t under circumstance n given by

$$P_{n,i,t}(\beta) = \frac{z_{avail,i,n,t} * e^{v_{i,n,t}}}{\Sigma_{j=1}^{\bar{J}} z_{avail,j,n,t} * e^{v_{j,n,t}}} \quad (18)$$

$V_{j,n,t}$ denotes the deterministic part of utility function without the error value of equation 11 excluding the error term $\varepsilon_{j,n,t}$ and $z_{avail,j,n,t}$ takes value 1 if alternative is available at time t during attack n else takes value 0. Equation 18 calculates the probabilities of MNL model for each factor individually. Each of the probability value of factors calculated is multiplied with the individual attack probability for the same instance, hence recognizing the recurrent attack nature in our data set. The probability is calculated when there we have multiple scenarios for each attack.

Estimation

Model estimation comprises of calculating the probability of each or alternative factor of, $j^*_{n,t}$ during time t of attack n i.e $Pj^*_{n,t}$ using equation18. The likelihood function of attack n with vector component β of the model is calculated by

$$L_n(\beta) = \prod_{t=1}^{Tn} P_{j^*_n,t} \quad (19)$$

Tn denotes number of attacks happened in different situations. For few data sets to calculate the resilience we include weights during the estimation of model in our calculation. Next phase and

last phase of the model is estimation after defining the model and parameters. The model is estimated using the utility function, calculated probability value considering the two possible cases of the values of each parameter.

For model estimation we use Hessian matrix, relying on numeric gradient. Hessian matrix is the partial derivative of function **f** and n variable (Braun,2017).

$$f: s^n \rightarrow s \tag{20}$$

Hessian matrix

$$H\ f(x) = \begin{matrix} S_{11}^2 & S_{12}^2 & S_{1n}^2 \\ : & : & : \\ S_{n1}^2 & S_{n2}^2 & S_{nn}^2 \end{matrix} \tag{21}$$

The model estimation of recoverability and vulnerability follows the same steps right from defining the parameters and model to calculating probability using Multi nominal logit and likelihood function and the end individual result of both are feed into MS excel to calculate the resilience of the smart grid using choice equation 1 of this research.

Nested Logit model

The nested logit model is applied to the dataset when utilities are further divided by parameters within the probability range. To calculate the resilience of the smart grid, we can use the Nested logit model along with the Multi nominal logit model for complex datasets with sub-factors being grouped into different levels, namely m to be top-level and q to be the lower level of the sub-

factors division. In nested logit model λ is used in calculating the probability of the model, unlike multi nominal logit at a given time t during attack n and factor i.

The same utility function of the MNL model can be used, the estimation of the model use the following probability values as equation 7

$$P_{i,n,t} = P_{m,n,t} P_{(o_m|m),n,t} P_{(i|o_m),n,t} \tag{22}$$

The λ values range between 0 and 1. This value can be changed depending upon the attack frequency and loss to the smart grid.

$$0 < \lambda_q \leq \lambda_m \leq \lambda_\gamma \leq 1 \tag{23}$$

We are normalizing the nested parameter to 1.

In a nested logit model, we design the tree structure with parameters of the utility function by grouping and sub-grouping.

This framework can be coded and implemented using various software with minor changes to the utility function or calculation of probability. In recent times the implementation of choice modeling has peaked in Python, R and certain packages have been written specifically to choice modeling.

CHAPTER V

LIMITATION AND CONCLUSION

This research strengthens the resilience of the smart grid with the proposed technique for better functionality of the smart grid after or before an attack. The goal of the proposed framework is to understand the possibility of the attack and its countermeasure using a choice modeling framework with multi-nominal and nominal logit. The limitation of this research is the mathematical model framework is not implemented in real-time; it's a conceptual model.

The mathematical model maximizes the restoration of the smart grid from a cyber-attack using this framework by predicting the output resulting in the necessity action/implementation to be made for future research or use.

Future research is required to convert the conceptual framework to the working model of the smart grid to enhance resilience. Future research will be done towards the working model.

REFERENCES

A. Hamlyn, "Computer Network Security Management and Authentication of Smart Grids Operations", *IEEE Power and Energy Society General Meeting — Conversion and Delivery of Electrical Energy in the 27st Century*, pp. 1-7, 2008.

Aloul, F., Al-Ali, A. R., Al-Dalky, R., Al-Mardini, M., & El-Hajj, W. (2012). Smart grid security: Threats, vulnerabilities and solutions. International Journal of Smart Grid and Clean Energy, 1(1), 1-6.

Alwosheel, A., van Cranenburgh, S. & Chorus, C. G. (2019). 'Computer says no'is not enough: Using prototypical examples to diagnose artificial neural networks for discrete choice analysis. Journal of Choice Modelling, 33, 100186.

Ashkrof, P., Homem de Almeida Correia, G., Cats, O., & van Arem, B. (2019). Impact of automated vehicles on travel mode preference for different trip purposes and distances. *Transportation Research Record*, *2673*(5), 607-616.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. In *Swiss transport research conference* (No. CONF).

Bierlaire, M. (2018, January 14). Introduction to Discrete Choice Models.

Retrieved from: https://courses.edx.org/courses/coursev1:EPFLx+DiscreteChoiceX+3T2017/course/

Biringer, B., Vugrin, E., & Warren, D. (2013). *Critical infrastructure system security and resiliency*. CRC press.

Braun, M. (2017). sparseHessianFD: An R Package for Estimating Sparse Hessian Matrices. *Journal of Statistical Software*, *82*(1), 1-22.

Calastri, C. (2020). Travel, social networks, and time use: modeling complex real-life behavior. In *Mapping the Travel Behavior Genome* (pp. 279-297). Elsevier.

Chen, X., & Chen, X. (2021). Data visualization in smart grid and low-carbon energy systems: A review. *International Transactions on Electrical Energy Systems*, e12889.

Dalziell, E.P., & McManus, S. T. (2004). Resilience, vulnerability, and adaptive capacity: implications for system performance.

Das, L., Munikoti, S., Natarajan, B., & Srinivasan, B. (2020). Measuring smart grid resilience: Methods, challenges, and opportunities. *Renewable and Sustainable Energy Reviews*, *130*, 109918.

D. P. Varodayan and G. X. Gao. Redundant metering for integrity with information-theoretic confidentiality. IEEE SmartGridComm'10, pages 345–349, 2010.

Farhangi, H. (2009). The path of the smart grid. *IEEE power and energy magazine*, *8*(1), 18-28.

Fang, X., Misra, S., Xue, G., & Yang, D. (2011). Smart grid—The new and improved power grid: A survey. *IEEE communications surveys & tutorials*, *14*(4), 944-980.

Fiksel, J. (2003). Designing resilient, sustainable systems. *Environmental science & technology*, *37*(23), 5330-5339.

FitzPatrick, J. (2011). NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 2.0.

Foreman, C., Ragade, R. K., & Graham, J. H. (2015). An Immersive Visualization Tool for Teaching and Simulation of Smart Grid Technologies. *arXiv preprint arXiv:1509.06293*.

G. A. Taylor, "Distributed Monitoring and Control of Future Power Systems via Grid Computing", *IEEE Power Engineering Society General Meeting*, 2006.

Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. *Computer networks*, *169*, 107094.

Hagenauer, J. & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. Expert Systems with Applications, 78, 273-282.

Han, Y., & Song, Y. H. (2002). Condition monitoring techniques for electrical equipment: A literature survey. *IEEE power engineering review*, *22*(9), 59-59.

Hess, S. & Palma, D. 2019; Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application, Journal of Choice Modelling, Volume 32, September 2019, 100170

Hoffman, E. (2007). Building a resilient business. *Raptor Networks Technology Inc*.

Hossain, N. U. I., Jaradat, R., Hosseini, S., Marufuzzaman, M., & Buchanan, R. K. (2019). A framework for modeling and assessing system resilience using a Bayesian network: A case study of an interdependent electrical infrastructure system. *International Journal of Critical Infrastructure Protection*, *25*, 62-83.

Ibne Hossain, N. U., Nagahi, M., Jaradat, R., Shah, C., Buchanan, R., & Hamilton, M. (2020). Modeling and assessing cyber resilience of smart grid using Bayesian network-based approach: a system of systems problem. *Journal of Computational Design and Engineering*, *7*(3), 352-366.

Khan, M. I. U., & Riaz, M. (2016). Various types of smart grid techniques: a review. *Int J of Multidiscip Sci and Eng*, *7*(8), 7.

Martinez, D., Henao, H., & Capolino, G. A. (2019, August). Overview of Condition Monitoring Systems for Power Distribution Grids. In *2019 IEEE 12th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives (SDEMPED)* (pp. 160-166). IEEE.

Matuszak, W. J., DiPippo, L., & Sun, Y. L. (2013, October). Cybersave: situational awareness visualization for cyber security of smart grid systems. In *Proceedings of the Tenth Workshop on Visualization for Cyber Security* (pp. 25-32).

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Mohammadian, A. & Miller, E. (2002). Nested Logit Models and Artificial Neural Networks for Predicting Household Automobile Choices: Comparison of Performance. Transportation Research Record: Journal of the Transportation Research Board, 1807, 92-100.

Mohammadali, A., Haghighi, M. S., Tadayon, M. H., & Mohammadi-Nodooshan, A. (2016). A novel identity-based key establishment method for advanced metering infrastructure in smart grid. *IEEE Transactions on Smart Grid*, *9*(4), 2834-2842.

Molloy, J., Schmid, B., Becker, F., & Axhausen, K. W. (2019). mixl: An open-source R package for estimating complex choice models on large datasets. Arbeitsberichte Verkehrs-und Raumplanung, 1408.

M. Stefan, J. G. Lopez, M. H. Andreasen and R. L. Olsen, "Visualization Techniques for Electrical Grid Smart Metering Data: A Survey," *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, 2017, pp. 165-171, doi: 10.1109/BigDataService.2017.26.

Newman, J. & Garrow, L. (2020). Stacked Hybrid Discrete Choice Models for Airline Itinerary Choice. Transportation Research Record, 2674(12), 243-253.

Nga, D. V., See, O. H., Xuen, C. Y., & Chee, L. L. (2012). Visualization techniques in smart grid. *Smart Grid and Renewable Energy*, *3*(03), 175.

S Hess, D Palma (February 17, 2020). Apollo User Manual version 0.1.0, Choice modelling center, University of Leeds.

S. McLaughlin, D. Podkuiko, and P. McDaniel. Energy theft in the advanced metering infrastructure. 4th Workshop on Critical Information Infrastructures Security, pages 176–187, 2009.

U.S Department of Energy, The Smart Grid, "n.d", https://www.smartgrid.gov/the_smart_grid/smart_grid.html

Van Cranenburgh, S., Wang, S., Vij, A., Pereira, F., & Walker, J. (2021). Choice modelling in the age of machine learning. *arXiv preprint arXiv:2101.11948*.

Venugopalan, S., & Rai, V. (2015). Topic based classification and pattern identification in patents. *Technological Forecasting and Social Change*, *94*, 236-250.

Yan, Y., Hu, R. Q., Das, S. K., Sharif, H., & Qian, Y. (2013). An efficient security protocol for advanced metering infrastructure in smart grid. *IEEE Network*, *27*(4), 64-71.

Yao, R. & Bekhor , S. (2020). Data-driven choice set generation and estimation of route choice models. Transportation Research Part C: Emerging Technologies, 121, 102832.

Yu, L., & Sun, B. (2012, July). Four types of typical discrete choice models: which are you using?. In *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics* (pp. 298-301). IEEE.

김발호. (2016). Analysis of the Impact of Cyber Attacks on Energy Management System in Smart Grid Environment.