ISBN: 978-628-95207-4-3. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2023.1.1.981

Graph Theoretical Modelling Experimental Prototypes to Support Cyber-security Investment Capabilities and Cost for Jamaican Firms

Claston Brown, MSc, Sean Thorpe, PhD
University of Technology, Jamaica, clastonbrown@yahoo.com, University of Technology, Jamaica, thorpe.sean@gmail.com

***Abstract*—This research paper posits the use of graph theoretical models to support the data visualization required for monitoring and measuring cyber-security investments for digital firms provided with sufficient capabilities assumed to manage the cost of such investments. This research present as vertices and edges within a graph that the total cost path function between vertices/nodes of a graph will be unique given that the total cost path value is dependent on the goal cost (i.e., the cost to enable the working environment) and the heuristic cost (i.e., the cost of an attack on the same working environment). This working environment are all represented as graph path cost and applies the principle of the travelling salesman problem within graph theory. Graph modeling supports the fact that we can use the nodes as points to track/determine the behavior of an investment as a function of time (t).**

***Keywords*—cybersecurity, investment, graph theory, technology, framework**

I. INTRODUCTION

The concept of representing cyber-security investments as a graph in the face of adversarial attacks is not new. The inspiration in this paper is drawn from [1].

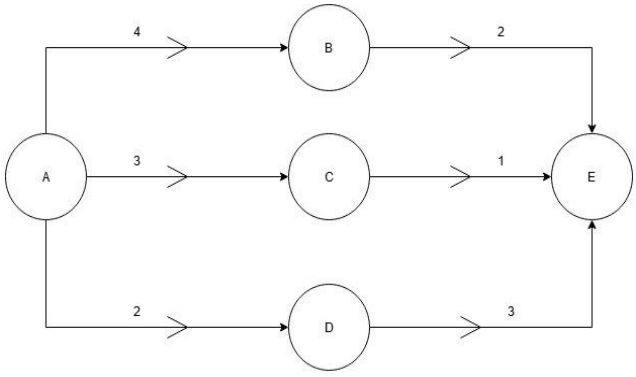

Figure 1.0 – The basic model of graph G with vertex v and Edge e

The basic data visualization of the nodes/vertex in Figure 1.0 including the edge E represents the data points over which the investment and system capabilities can be calculated to show what Jamaican firms are potentially facing when cost changes within the organization are influenced by attacks. In this paper the total cost path f(n) represents the cost of a cyber-security investment.

Digital Object Identifier: (only for full papers, inserted by LACCEI).
ISSN, ISBN: (to be inserted by LACCEI).
DO NOT REMOVE

The points between vertices represent your goal cost (g(n)) which denotes the cost of enabling the system capabilities to avert an attack. This could mean making sure you hired the right talent, having a resilient infrastructure in terms of hardware, software, governance and compliance policies. On the graph edge itself E you may now have hidden cost which we call the heuristic cost (h(n)). This h(n) cost could be a zero-day ransomware attack or any form of an advance persistent threat not yet seen. In Figure 1.0 above the weighted values in the graph denote the g(n) cost and the assumption in this example is that the h(n) = 0. Hence where f(n) = g(n) + h(n) , the total cost path cost function goes down when the f(n) and h(n) goes down and the reciprocal is true. The discussion below presents standard underlined computational properties of Graph G where we treat this graph as a Cyber-security investment graph function. So, the properties are as follows.

(i) Graphs can be directed or called digraphs [2]. In essence the movement of the function f(n) is unidirectional, which means the cost path function does not change and also suggests that the cyber-investment cost does not also change.

(ii) Graphs can be undirected, which means the f(n) function is bidirectional, to the graph origin around a point. In essence your cyber-security investment can be affected positively or negatively in the event that the graph direction changes. This is to suggest that the cyber-investment can be positive or negative and in essence takes the absolute dollar value of the investment and not its relative value as suggested in Figure 2.0. Again, the weighted values in the graph here could be treated as the g(n) values and the h(n) value could either be 0 or randomly generated. The total cost path function is therefore dependent on the calculated values of the g(n) and h(n) values together.

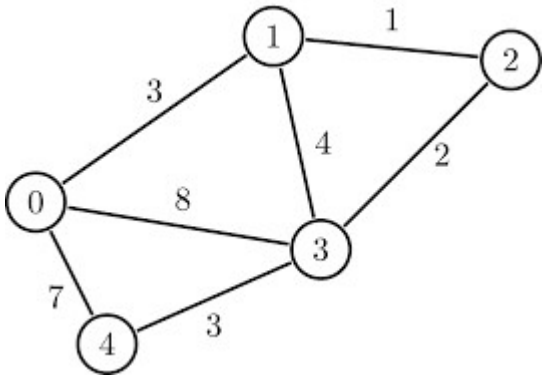

Figure 2.0 - Bi-directional (undirected cyber-investment graph)

(iii) Graphs can be cyclic to suggest that the starting vertex and the ending vertex are the same nodes within a graph, such that wherever your cyber-security investment started off it can end up back at the same place. So where an investment node vertex starts at A, it means sometimes regardless of the dollar investment to support the enabled capabilities, eroded profits due to consistent attacks, can create cycles in spending such that the starting point of the investment also represents the ending point. And that is to suggest that there may be no long-term net positive return on investment, hence you have a loop back on the investment itself for the project of the Jamaican company.

So as you can see from Figure 3.0, there is a cyclic loop on vertex/node A, B, and C. In each case for a cyber-security investment, what is depicted is a financial loop on the investment, which starts and end with the initial cyber-investment as the final investment outcome within this given context.

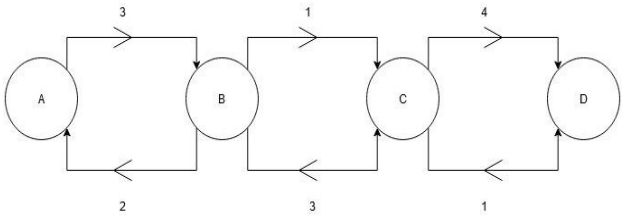

Figure 3.0 - Directed and cyclic cyber-investment graph

(iv) A graph can be acyclic graph which is to suggest that the start and end vertex are different, and hence is to suggest that what you started off with as an cyber-security investment will not be the same value at the end as seen in Figure 4.0 below. The reasons for this could be that where the cost of an adversarial attack adds to the dollar cost of an investment, it means with time the outcomes of the cyber investment starting at vertex A now ends at vertex E.

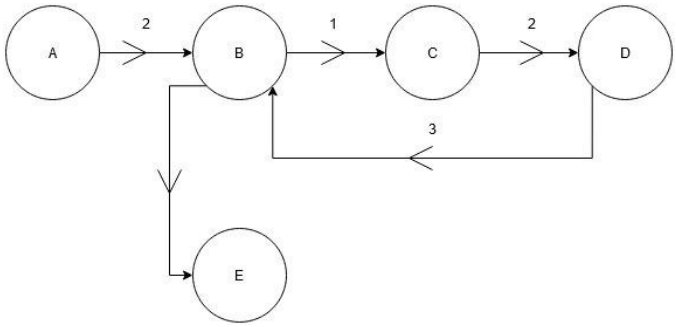

Figure 4.0 An acyclic cyber-investment graph

(v) Graphs can be Hamiltonian – in that no vertices traversed between the start and end vertex through the graph traversal from the start vertex [3]. For example, let's say this type of graph investment is to suggest that where you started at point A with the investment function, all nodes are traversed once (not repeated), and hence produces a different cyber-investment cost every time the function is computed.

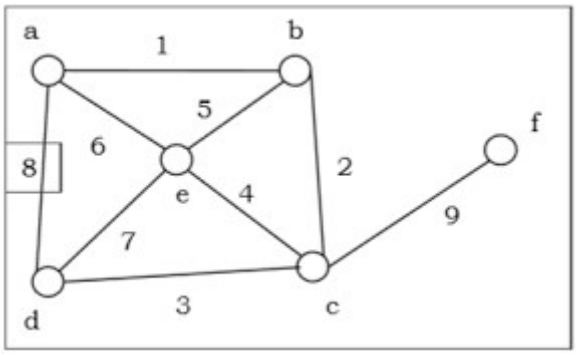

Figure 5.0 – Hamiltonian cyber-investment graph

(vi) Graphs can be Euler – that is a graph where one node/vertex is repeated in traversal between the start and end vertex. Within the context of the cyber-security investment, this means where a node can be repeated or revisited once then this suggest that you can have an instance where the dollar value of an investment could repeat around a node point. Hence ebbs in spending can have patterns such that highs or lows may show repeated patterns of the spending particularly due to some cyber-attack. This may be suggestive that the company may not have taken adequate steps to harden the network environment, and hence the repetition of attack sequences finds us in a situation where the repetition at some point – forces a previous pattern of spend which should have been avoided. Such circumstances show up or highlight the situation of the Jamaican company not paying attention to key metrices that will seek to safeguard the network perimeter with time that may be forcing such spend. Notwithstanding zero day ransomware attacks, where companies were never prepared for the emergent Advance Persistent Threats (APT) within the environment could lead to these abstract and repeated graph patterns.

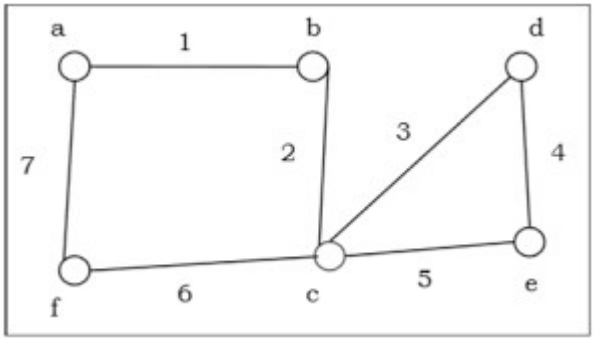

Figure 6.0 – Euler cyber-investment graph

(vii) Graphs can be directed, cyclic and Hamiltonian, and they can also be acyclic and Hamiltonian, and directed. Graphs can be directed cyclic and Euler, they can also be acyclic directed and Euler. These properties or attributes are mix of the pointers already raised above as seen in the individual properties. This context seeks to hybrid the features to show the range of multiple attributes the graphs can have, and hence shows the various pathways that a cyber-security function can have, given the variables of the g(n) and h(n) elements that determine the total cost path of the investment graph.

(viii) Graphs can also be bipartite in nature where, you can have traversals around a particular vertex or node within a graph. It is sometimes described as a vertex loop back. In the principle of the firm, the suggestion here is that the cyber-security investment function may be zero, or a lack of an investment against cyber-attack and as such you have with time, the same impact on the network nodes of your system infrastructure.

From Figure 7.0 we see that for a particular cyber-security investment, that on node B, the graph could halt, in that the investment in the overall firm is now experiencing significant down time on network node B which may be a particular function or operation within the company that seems to have an investment that is not getting anywhere.

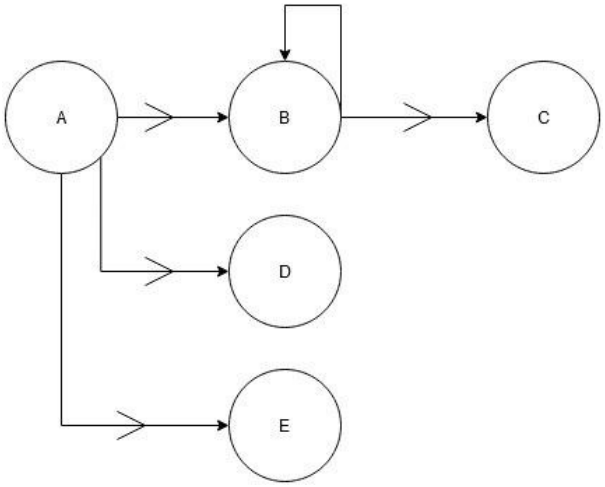

Figure 7.0 - A directed Acyclic bi-partite cyber-investment graph.

In addition to the above features that can be used to visualize the cyber-security investment and capabilities matrix for the Jamaican firm, its visualization provides rich context that if modelled into a data dashboard for the Chief Executive Officer (CEO) of the firm it would provide a useful contribution with respect of the CEO and Chief Information Security Officer (CISO) being able to track the investment spending and capabilities of the organization.

II. Discussion

So, from the above graph visualizations we can in fact fashion case examples of how Jamaican businesses fit into this situation. This research actually seeks to use the UK breach survey studies in [4] to make the comparison modeling in the absence of the data needed from Jamaican companies, however the model provided by the graphs will allow us to make a clear data visualized dashboard of how we should treat these cost comparisons on cyber-security investments and the enabling capabilities. If we take the discussion one step further once we have established these graph properties, our next step is to be able to look at the computational cost involved in searching these graphs. These search costs, as previously stated, represent the f(n) value to the digital firm. This research posits as a new contribution to the literature that these graphs as computational graphs for tracking cyber-security investments should be divided into informed search graphs and uninformed search graphs. An informed search graph assumes that it finds the most optimal cost path to compute cost of cyber-security investments. In the case of an informed cost search cyber-security investment graph, the heuristic cost works out over the long run to be lower. And hence when this research paper refers to this heuristic cost it is speaking of the cost of adversarial attacks on the company's assets. While an uninformed search graph assumes that the heuristics tend to work out to be way higher, given the spate of adversarial attacks or advance persistent threats which could not have been predicted. Against the above background, what our research now posits is the need to demonstrate these behaviors of the graphs i.e., informed, and uninformed searches as a direct relation to cybersecurity investment. Very specifically for an informed search graph - the known Algorithm that sufficiently tracks this behavior is the A* search algorithm, where the total f(n) cost on the graph is computed and the optimal path cost function returned over the range of the graph. That optimal cost path function represents for us the tracking of the lowest cost to the cyber-security investments considering the various other path cost alternatives available to us in the graph. Our research adopts this strategy from the well-established travelling salesman problem discussions from Discrete mathematics and the analysis of algorithms within the computing literature.

Following on from the A* search algorithm measure, a semi-informed search graph option that comes next is the best first search graph or what is sometimes referred to as the K-nearest neighbor (knn) option. This graph computes the nodes from the start to end, and traverses from the start vertex the nearest neighbor with the lowest heuristic value function and then take that path. The clear danger here is that the heuristic costs are unpredictable, and as such it could be higher than expected. However, on the flip side it could be lower than expected and hence what obtains if it is lower than expected the best possible outcome on such a graph is that it becomes A* search cost optimal, and in the worst case, if the heuristic cost keeps monotonically increasing is that the graph retards to a greedy search cyber investment, where this is the poorest result of a cyber-security investment. While the authors in [4] demonstrate a case of how the KNN to track cybersecurity investment spending based on a set of trend reports of data breaches within the United Kingdom between 2018 and 2019, a clearer understand on the graph traversal paths that influence the spend was not well understood and neither could the researchers visualize the total cost path function to the UK firms in that study. The study however motivates the argument that if we apply these graph models we can allow for the firms to better understand the cyber-risk and spending specific to a breach. To further highlight, if the company or firm has been breached, we can use graph coloring techniques on the graph paths to demonstrate that breach has occurred. In other words the highlighted or colored graph paths indicate a cyber-investment total cost path function which is high.

III. Experimental Analysis

To support the design of the theoretical graphs so far discussed in this paper, our approach is to model an Application programming Interface (API) that can simulate these graph

behavior as a visualized cyber-investment function for the CEO of the Jamaican business. We encapsulate the graphs using tools like Power Bi and Java to capture the design of the graph properties. We set this up on a dedicated Windows 10 machine. While the graph modeling is done, we input test functions that calculate g(n) and h(n) values. The h(n) values are randomly generated. For graph paths that are expensive we apply the graph coloring technique to track and determine the elements of high risk or sources of potential breach to the organization. Please note for this experimental design study, the data values represent synthetic data just to support the simulation. We compute the A*, KNN and greedy search behavior into the graph to show the basic design elements of our working prototype. Although our prototype is still in its early stages, the diagrams below support a mockup of the same.

So from a basic investment prototype design looks like this:

Screen Example.

Enter the start Node ---------- End Node ----------

Calculate the total cost path --------

Output of the Visualized graph with the weighted total cost investment Function

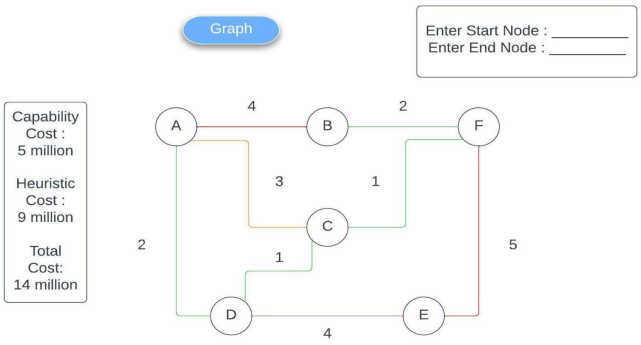

Figure 8.0 - Output cyber- security risk investment graph based on node input data.

We assume that the start node represents an enabling infrastructure in the Jamaican company network, and the weighted values for several such nodes calculates the total cost path function. When the CEO who is your test user clicks on any of these nodes it generates a further analysis of the graph and its individual nodes to provide a detailed description of the total cost function parameters of the investment. These cost components capture the single and annual loss expectancies of the digital firm, the return on Investment and the exposed risk to the firm. The variables of the cost calculation are extracted using the Delphi method to provide the empirical cost analysis.

For any nodes selected on the graph in G in figure 8.0 above or even the selection of a single node, it generates the prototype output similar to the visualized results below.

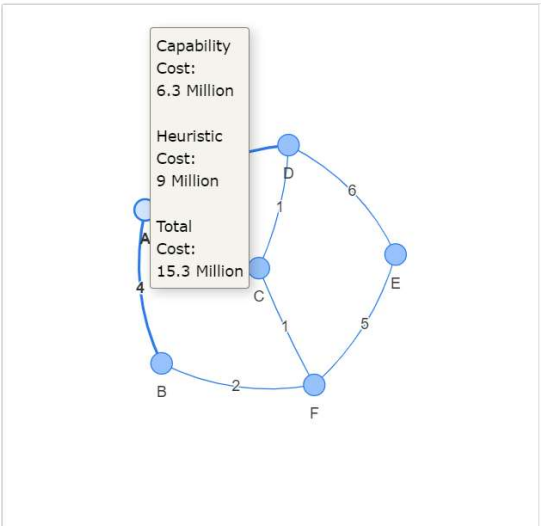

Figure 9.0 - Output empirical analysis of the total cost function given in the cyber-investment risk graph from Figure 8.0.

IV. Materials and Methods

We analyze our design prototype in relation to the end users, who include CEOs, CISOs, and Chief Information Officer (CIO) who are analyzing the application. Our software prototype is intended to track/measure how the CEO, CISO, and CIO may evaluate the level of cyber-security investment based on system capabilities and the changes that may occur when the system is exposed to an attack. Our research aimed to solve two (2) research questions:

i. Using annotated data graphs and a human computer interaction framework, can you quantify and display the level of cyber-security investment within a corporation given their cyber-security capability?

ii. How effective would this methodology be at tracking and visualizing levels of cybersecurity investment within the organization?

A. Methodology

This section outlines our approach to answering the research questions. This study is classified as a mixed technique exploratory study. The reason for the mixed methodology is that the study began as a complete quantitative experimental study in which we modeled an actual software visualization of a prototype in Python that represents the cyber-security spend based on the capabilities and heuristic cost of the Chief Executive Officer. However, in evaluating the prototype's usability, when our population was 100, our sample size was 15, and only 8 replies were ultimately used to support the final review. In qualifying the quantitative results, the eight outcomes were not statistically significant. As a result, we used the eight (8) responses to support a thematic analysis based on the qualifying responses' description and interpretation. This last point backs up the qualitative method. As a result, against this backdrop, the entire study is benchmarked by a mixed approach as the final output, where we map the visualization of

an organization's cyber-security investment spending as the concept [5].

Our findings in modeling the annotated graphs were driven by a set of toy experiment simulation designs in which we model the graph nodes as actual locations within a basic company network, and the edges between these nodes are weighted to support the cost of a cyber-attack or what we call our heuristic function within the graph, while the nodes themselves represent the capability cost or goal cost function within the graph. The study's approach was to evaluate appreciation from a group of end users of Chief Executive Officers (whose identities will be kept private). We targeted fifteen (15) CEOs, CISOs, and CIOs whose general competence was not Information Technology based on background from a Cyber-security focus group of 85 individuals on an instant messenger platform, managed by the authors as part of a convenient sampling. This type of sampling was carried out in order to acquire an unbiased sampling of our data. In the grand scheme of things, this would allow for a more representative selection of Jamaica's roughly 400 small to medium-sized firms that may not be core technologists by training.

In a global world, the idea is that more CEOs, CISOs, and CIOs are not technologists by background as members of the end user community that we have targeted in this usability study to measure the usefulness of our software design prototype designed to track the data visualization of cyber-security investment and dynamic capabilities [6],[7]. We shared a zoom recording of the mocked-up models of the software prototype with them on Google Drive and allowed them to submit asynchronous comments based on the shared prototype.

Our specific questions to the CEOs, CISOs, and CIOs about the software prototype were as follows: (i) How useful do you find our software prototype design concept? Please rate it on a scale of 1 to 5, with 5 being very good and 1 being poor. (ii) What general recommendations would you make for the software prototype improvement. We only obtained responses from eight (8) of the fifteen (15) individuals who were identified. The responses to the questions are shown below.

B. Results

According to the statistical responses, just one (1) of the eight (8) responses rates the tool's acceptance as (1 out of 5 or a 20% acceptance rate). Based on the designed prototype, the average acceptance percentage of our software prototype is 75% as outlined in Figure 10.0.

In terms of thematic reviews, what we gleaned from the list of recommendations from all eight (8) responses to our usability survey is that, while the front end user interface represented a good idea, what they would have also liked to see is the back-end interfaces with respect to how the data was collected with respect to capabilities cost and the heuristic cost based on attack vectors which the nodes could encounter.

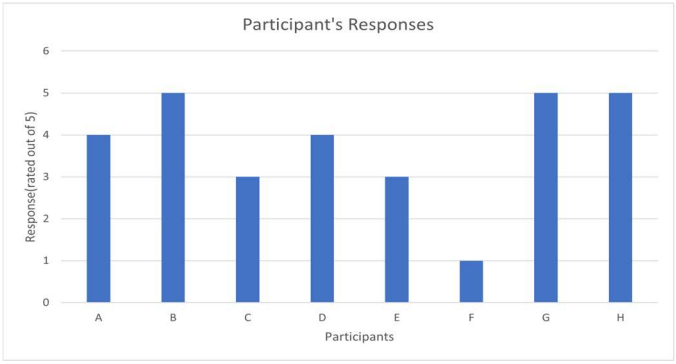

Figure 10.0 – Participants survey response to the effectiveness and usability of the prototype.

There was also a need to further clarify a node as a virtual versus a physical center, as well as the reality that where most data aggregation today occurs within a cloud computing environment, the real cost would be the nodes interpreted as threat vectors, with a need for further refinement of the threat types that can impact the node, as well as the ability to disaggregate the node profile based on threat types and provide the heuristic cost based on those threat types. When a threat type is reported at no cost, the CEO, CISO, and CIO should be alerted to the requirement to establish a reasonable spend against the mitigation of that specific type of assault.

The summary assessment of our usability experiments tied to the basic software prototype designs that we had put together with these annotated graphs is that the feedback was favorable about the design's suitability based on the system characteristics, the user characteristics, the task specific context and its characteristics being that we were specifically looking at the cyber-security investment capabilities and heuristic cost related expenditure. The study also found that the intended and actual outcomes were well quantified, indicating that our user design was properly integrated into the Human Computer Interaction framework alluded to in this study.

In conclusion, our research has provided us with a comprehensive usability experimental analysis based on our iterative throwaway prototyping model within the context of a Human Computer Interaction framework of annotated data graph models to build out the concept of cyber-security investments related to capability and heuristic cost spend in the face of advanced persistent threats that can impact your company's network. The revelation is that our modeling technique is not restricted to developing self-sufficient prototype apps like dashboard applications for small and medium-sized businesses, but also for extremely large businesses.

C. Limitation

While our usability experiments supported approximately 15 CEOs, CISOs, and CIOs as a reference to the convenient sampling, a more representative sampling within a large scale of enterprises across multiple sectors and industries was not available to support what we would have liked to implement this overall study given that larger teams of researchers would have been required to drive this outcome.

V. Conclusion

Our research has provided us with a comprehensive usability experimental analysis based on our iterative throwaway prototyping model within the context of a Human Computer Interaction framework of annotated data graph models to build out the concept of cyber-security investments related to capability and heuristic cost spend in the face of advanced persistent threats that can impact your company's network. The revelation is that our modeling technique is not restricted to developing self-sufficient prototype apps like dashboard applications for small and medium-sized businesses, but also for extremely large businesses.

In conclusion, we have clearly answered both research questions. (i) We visualized and tracked the intentions of a cyber-security spend based on the capabilities and heuristics given the various attack vectors that a firm may face using the foundations of graph theory and the various assumptions and properties within these graphs [8],[9]. By building a semi-functional prototype using simulation diagrams and applying programmatic coding with Python as a reference to the design interfaces, we were able to use the annotated data graphs to display the cost functional behavior of a spend for the firm. For tracking the system characteristics of our prototype based on its functional specifications, we placed it within a layered Human Computer Interaction framework [10]. We were able to map the user characteristics related to the end user needs (in this case, the CEO, CIO, and CISO of the firm) by (ii) evaluating the concerns within the first research question through a series of corroborating usability experiments against our semi-functional prototype that tracked that user's intentions and behavior based on the expected and actual outcomes observed from our user interface design. As our HCI method, we used rapid software prototyping. Our work makes a significant contribution to the fields of Human Computer Interaction, Cyber-Security, and core computer science as a result of how we can operate to drive Information Systems environments, specifically Enterprise Security management environments for small, medium, and large businesses.

We learn from the study and the numerous recommendations that a lot more effort is still needed to look at these cost factors inside the graph as continuous variables rather than static cost variables, as we did in our research. In terms of the continuous variable function, we must recognize and appreciate that while the firm's capability cost in terms of things like policy change, governance, infrastructure, and human capacity may have some level of stability, in the real world based on the size and operation of the firms across various industries, the capabilities are dynamic, and thus the firms would need to show through their accounting procedures and general practices.

References

[1] Behavioral and Game theoretic investments , https://engineering.purdue.edu/dcsl/publications/papers/2020/game-theory-security_tcns20.pdf, retrieved February 3 2023 .

[2] Johnsonbaugh, R. (2018). *Discrete mathematics.* Pearson.

[3] Dawood , H. A. (2014). Graph Theory and Cyber Security. 3rd International Conference on Advanced Computer Science Applications and Technologies (pp. 90-96). IEEE Xplore.

[4] De Arroyabe, I. F., Arranz, C. F., Arroyabe, M. F., & de Arroyabe, J. F. (2022). Cybersecurity capabilities and cyber-attacks as drivers of investment in cybersecurity systems: A UK survey for 2018 and 2019. *Computers & Security*, 1-17.

[5] Dor, D. and Elovici, Y. (2016). A model of the information security investment decision-making process. *Computers and Security*, 63, 1-13.

[6] Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350.

[7] Teece, D. J. (2017). Dynamic capabilities as (workable) management systems theory. *Journal of Management & Organization*, 24 (3), 359–368.

[8] Sadavare, A. B., & Kulkarni, R. V. (2012). A Review of Application of Graph Theory for Network. *International Journal of Computer Science and Information Technologies*, 5296-5300.

[9] Zhang, W., Chien, J., Yong, J., & Kuang, R. (2017). Network-based Machine Learning and Graph Theory Algorithms for Precision Oncology. *Nature Partner Journals Precision Oncology*, 1-15.

[10] Zhang, P., & Li, N. (2005). The Intellectual development of human-computer interaction research: A critical assessment of the MIS Literature (1990-2002). *Journal of the Association for Information Systems*, 6 (11), 227-292.