Problem Statement

A Technological World

Technology is a powerful agent of change that has shifted the focus of corporations and individuals from face-to-face interactions to online communication. Web applications deployed on accessible backend services such as Heroku or AWS can link family members who live in different parts of the world or companies trying to reach an international consumer base. On top of that, in the public sector, shared hospital networks (usually hosted on cloud services) can help doctors receive real-time electronic health records. The comprehensive streaming of data present in most hospitals can improve their ability to take action regarding critical patient data.

Simply put, technology represents convenience and accessibility in the eyes of individuals, and humans have incorporated this new tool into most aspects of their lives. For example, in the United States alone, Internet traffic for banks grew by 77.6% between July 2000 and July 2001 [?]. On top of that, online retail sales have increased to 269 billion USD in 2004 from a mere 45 billion USD in 2000 [?]. Finally, since the healthcare industry is under constant pressure to reduce costs and improve patient care, information technology can help a hospital manage risks and improve organizational performance [?]. Technology such as computer tomography (CT) scanners and X-ray machines help doctors identify issues much faster. So, in the 21st century, hospitals and businesses alike have seen a growth in users, highlighting the ever growing impact of technology.

Uniqueness of This Researcher's Current Work

The new research performed addresses the flaws in previous research and allows for greater theoretical optimization of the knapsack problem. In this new algorithm, Quantum Save, the higher-order quantum genetic algorithm is implemented on the IBM Q32 simulator which allows the algorithm to better simulate a quantum state. Quantum Save uses the increased complexity of the Q32 simulator to better mimic the random variations that occur in a genetic algorithm. In addition to that, although Kucharski and Nowotniak [?] provided a range for the amount of probability amplitude manipulation, they didn't provide a certain value. So, Quantum Save will consist of tuned hyper parameters for which the optimization of the Knapsack problem is the greatest. Also, with Quantum Save, a disaster condition will be added to the higher-order QGA to prevent premature convergence. In fact, social disasters technique, which applies a catastrophic operator once a local maximum is reached is one of the best methods to obtain the global extremity point instead of the local maxima [?]. In Quantum save, the disaster condition will implement a simple quantum catastrophe operator. Therefore, by tuning the higher-ordered QIGA, the probability of reaching a global maximum will increase drastically, and the optimization value for the knapsack problem will be much higher.

In addition to that, Quantum Save can help hospitals and defense contractors save millions by allocating the budgets of small and medium sized targets. In previous research, the cyber security investment model was based upon simulated values and game theory. And with the existing optimization models, like the one proposed by Rakes et al. [?], the model performs significantly better for bigger targets since there is more budget to work with. The model does not face the full effect of the budget constraint and can easily choose countermeasures to maximize mitigation. However, it is the smaller-medium sized targets who are the most vulnerable to malware from adversaries. So, Quantum Save uses threat factors and mitigation values from real-life case studies to improve the budget allocation of these small-medium sized targets. Specifically, the data set compiled by Rakes et al. [?] in Figure 3 is inputted into the quantum algorithm. Researching this topic will allow the scientific and engineering community to accomplish countless tasks that have a limitation factor and an optimization value. Quantum Save can also improve the cybersecurity defenses of hospitals and defense contractors.

Investigation/Research

In this research problem, the extent to which Quantum Save can optimize the spending of a cybersecurity budget is investigated. Using the data set defined by Rakes et al. [?], Quantum Save will be implemented and an optimal investment strategy can be determined. The project will consist of two phases where the first phase involves tuning the Quantum Save algorithm with the optimal probability amplitude manipulation and implementation of the disaster algorithm. In this phase, it was hypothesized that the probability amplitude manipulation along with the disaster algorithm implemented on a POWER 9 architecture would result in the optimization of the Knapsack problem. For the first phase, the independent variable is the manipulation of the probability amplitudes and the implementation of the disaster algorithm while the dependent variable is the optimization of the Knapsack problem.

The second phase of the project involves applying the Quantum Save algorithm to the cyber security modeled previously defined and see if this algorithm can provide an optimal investment strategy for CISO's. In this phase, it was hypothesized that if the Quantum Save algorithm is implemented in the cyber security investment problem, then the investment strategy that is obtained will be better when compared to previous case studies (specifically, the linear optimization model by Rakes et al. [?]). This is because Quantum Save has a better optimization of the knapsack problem when compared to that of previous research. Also, quantum computing would allow for a better simulation of random variability in the genetic algorithm. In this phase, the independent variable is the implementation of the QIGA algorithm with a cyber security purpose while the dependent variable is the investment strategy that is obtained after implementing the QIGA algorithm.

Quantum Save consists of six major steps and these individual steps are performed multiple times throughout the evolution process. Every time these steps are performed is called a generation and through multiple generations, the population should evolve to produce a greater optimization of the knapsack function. The steps in each generation derive from Darwin's theory of evolution which states that evolution is a cause of overpopulation, competition between individuals of the species, and eventual survival of the fittest. However, before implementing the QIGA-2 algorithm on the IBM Q32 computer, various libraries including Qiskit by IBM and NumPy must be imported in order to connect with the quantum computer.

Procedures

Initializing the Population

The first section of the quantum genetic algorithm is reliant on the idea of overpopulation and natural variability in genes. As Darwin concluded, evolution is spurred when there are too few resources for too many individuals and random variability in the population cause some individuals to be better than others. In the first step, the quantum population is created and initialized. In this case, the population is an array of 32 qubits, the maximum number of qubits available on the IBM Q32, and each quantum chromosome represents a row in that array

Algorithm 1 Quantum Save – Phase 1

```
$n \leftarrow$ number of countermeasures
$b \leftarrow$ budget
$probabilityAmplitude \leftarrow [0.5, 0.5, 0.5, 0.5]$
$generation \leftarrow 0$
while $generation < 10$ do
    $quantumRegister \leftarrow 0$
    while $quantumRegister < 4$ do
        $classicalRegister \leftarrow$ classicalRegister[n]
        $quantumCircuit \leftarrow$ classicalRegister + quantumRegister
        $quantumCircuit \leftarrow$ probabilityAmplitude
        $job \leftarrow measure(quantumCircuit)$
    $quantumRegister{+}{+}$
```

The array of 32 qubits is arranged to simulate a Think-Tank of four Chief Security officers. The goal of the Think Tank is to create a recommendation list of 8 countermeasures which defense contractors and small-medium targets across the world can implement in order to protect themselves from foreign attacks. Each of the four rows in the qubit array, or matrix, represents a Chief Security Officer in the think-tank, and each of the eight columns in the qubit array represents a countermeasure on the recommendation list. If the qubit returns a value of 1, then the countermeasure is added to the recommendation list. On the other hand, if the qubit returns a value of 0, then the countermeasure is removed from the recommendation list.

After the quantum population is created, each qubit register in that population is initialized with a linear superposition of states. By starting out with an equal chance of getting one and zero, there is natural variability in the population and the genetic algorithm can be more effective. In the case of the cyber security application, the linear superposition of states that the quantum population is initialized with represents an equal probability of a defender implementing a certain control and not implementing that control. Therefore, there is natural variability in the population and the genetic algorithm is effective.

Finding an Optimal Plan

Now that the population has been created and has been initialized with a linear superposition, the second section of the quantum genetic algorithm can begin. This section the individuals with the best random attributes are chosen and the fittest individuals in the population, or those who pack the knapsack the best, will remain for the next generation. After the initialization of the quantum population, it will be classically measured by collapsing the existing wave function. Depending on the alpha and beta values present in each qubit register, a binary string of size two, either 00, 01, 10 or 11, will be obtained from each register

Algorithm 2 Quantum Save – Phase 2

```
$CISO \leftarrow 0$
while $CISO < 4$ do
    $binaryString \leftarrow$ ""
    $result \leftarrow$ ""
    while $result$ in job do
        $binaryString \leftarrow binaryString +$ Most Frequent String
    $countermeasure \leftarrow 0$
    $knapsack \leftarrow 0$
    while $countermeasure$ in binaryString do
        $knapsack \leftarrow knapsack + countermeasure$
        if $knapsack > b$ then
            $knapsack \leftarrow knapsack - countermeasure$
        $profit \leftarrow$ amount saved[$countermeasure$]
    $bestIndividual \leftarrow max(countermeasure)$
```

During each fitness evaluation, or the conversion between quantum and classical states, the qubit register is measured 1024 times and the most consistent binary string is returned. The binary strings from every four registers will then be combined until four classical strings are obtained, one for each individual in the Think Tank.

After quantum measurement, using the Copenhagen interpretation, is complete, the binary strings are evaluated in order to determine the best individual. Depending on the binary string, which essentially represents the controls that are implemented by the defender, a certain mitigation will be received. In this case, if there is a 1 that means that the defender chose to implement a specific control. On the other hand, if there is a zero, then the defender didn't implement a certain control. Apart from the profit value, a total weight for the knapsack was also obtained, or the total cost of implementing all of the controls, and that total cost was compared to the cost cap, or the budget, set before the first generation. After each binary string is evaluated, the binary string, and respective individual, with the highest profit value is stored as the best chromosome. With the application to a cyber security investment problem, the best chromosome will be the defender that implemented the controls that resulted in the greatest mitigation of the attack.

Amplitude Manipulation

Since the best strategy to mitigate an attack on the defender's small-medium enterprise has been determined, the probability amplitudes of each qubit register will be manipulated in the following manner.

Algorithm 3 Quantum Save – Phase 3

```
$binaryArray \leftarrow binaryString.split()$
$CISO \leftarrow 0$
while $CISO < 4$ do
    $countermeasure \leftarrow 0$
    while $countermeasure$ in $binaryArray$ do
        if countermeasure != bestIndividual then
            $probabilityAmplitude \leftarrow$ manipulate[$probabilityAmplitude$]
        probabilityAmplitude $\leftarrow probabilityAmplitude -$
```

Initially, the first binary pair in the best individual will be obtained and depending on the value of that pair, the probability distribution will be normalized to one. On the other hand, all of the other probability amplitudes will decrease by a factor of *probabilityAmplitude*, a hyperparameter in the algorithm. In this case, the probability of a defender implementing the successful controls increases while the probability of a defender implementing the unsuccessful controls decreases. The factor of probability manipulation, *probabilityAmplitude*, is a hyper parameter of this al- gorithm and this hyper parameter was tuned by testing out different values for and observing the optimization of the knapsack problem. In the end, the controls that resulted in higher mitigation of an attack were encouraged while controls that resulted in lower mitigation of an attack were discouraged.

At this point, the algorithm has completed one generation of fitness evaluations. In order to obtain an accurate optimization of the knapsack problem, this process needs to be repeated for 10 generations in order to get a higher optimization of the knapsack problem. At the end of these ten generations, the data of final profit outputted by Quantum Save will be obtained and that data can be compared to that of previous research. However, now that the probability amplitude manipulation is complete and the optimal value has been determined, a disaster algorithm will be implemented. In this case, the disaster algorithm will simply reset the population to a linear superposition of states in the middle of Quantum Save. However, the generation that the disaster algorithm is implemented, is a hyper parameter of this algorithm and this hyper parameter was tuned by testing out different generations. In the end, the location of the disaster algorithm was chosen by the overall mitigation of an attack by the defender.

After one instance of the algorithm, or 10 generations, has completed its run-time, the data of final profit outputted by Quantum Save will be obtained and that data will be compared to that of previous research. On top of that, the different hyper parameters including the value and the generation for the disaster algorithm will be tuned.

Analysis & Testing

The first phase of data collection involves tuning the actual hyper parameters of the quantum genetic algorithm. One hyper parameter that needs to be tuned is the value that describes how much the probability amplitudes are manipulated when the quantum population is evolving. This hyper parameter is important because it is fundamental when describing the probability amplitude manipulation that occurs in each generation of the genetic algorithm. It helps determine which cyber security controls are optimal and successful in mitigating attacks. To determine which hyper-parameter obtains a higher optimization of the Knapsack problem, an ANOVA test and a two-sample t-test will be utilized.

The second phase of the project involves the application of this algorithm to cybersecurity. In order to understand the relationship between the output of Quantum Save when compared to the output of previous research, the distribution of Quantum Save's output at each input range will be determined. Depending on the distribution of outputs when compared to previous research, conclusions will be drawn regarding the usefulness of the algorithm in allocating the budgets of small cybersecurity companies. In addition to that, the run time of Quantum Save will be determined by analyzing the experimental data surrounding the run time at different countermeasure recommendation list lengths. After finding a pattern in the experimental data, a theoretical analysis of the algorithm will be performed by analyzing the Pseudo code.

References

[1] "The Copenhagen and Many Worlds Interpretations of Quantum Mechanics." [Online]. Available: http://afriedman.org/AndysWebPage/BSJ/CopenhagenManyWorlds.html

[2] "Deloitte NASCIO Cybersecurity Survey | Deloitte Insights." [Online]. Available: https://www2.deloitte.com/us/en/insights/industry/public-sector/nascio-survey-government-cybersecurity-strategies.html

[3] A. Birkett, "How to Get Over Local Maximum? Everyone Hits It Eventually." [Online]. Available: https://cxl.com/blog/local-maximum/

[4] R. Bradley, T. Byrd, J. Pridmore, E. Thrasher, and R. Pratt, "An Empirical Examination of Antecedents and Consequences of Its Governance in US Hospitals," *Journal of Information Technology*, vol. 27, 2012.

[5] W. L. Chapman, J. Rozenblit, and A. T. Bahill, "System design is an NP-complete problem," *Systems Engineering*, vol. 4, no. 3, pp. 222–229, 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sys.1018

[6] A. Fielder, E. Panaousis, P. Malacaria, C. Hankin, and F. Smeraldi, "Decision support approaches for cyber security investment," *Decision Support Systems*, vol. 86, pp. 13–23, Jun. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923616300239

[7] R. Forrester, "The Quantum Measurement Problem: Collapse of the Wave Function Explained," *Social Science Research Network*.

[8] A. Garg and S. Goyal, "Vector Sparse Representation of Color Image Using Quaternion Matrix Analysis based on Genetic Algorithm," 2017.

[9] S. Gootman, "OPM Hack: The Most Dangerous Threat to the Federal Government Today," *Journal of Applied Security Research*, vol. 11, pp. 517–525, 2016.

[10] H. Mia, M. A. Rahman, and M. Uddin, "E-Banking: Evolution, Status and Prospect," *Journal of ICMAB*, vol. 35, 2007.

[11] Y. Mirsky, T. Mahler, I. Shelef, and Y. Elovici, "CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning," *arXiv:1901.03597 [cs]*, Jun. 2019, arXiv: 1901.03597. [Online]. Available: http://arxiv.org/abs/1901.03597

[12] A. Montanaro, "Quantum algorithms: an overview," *npj Quantum Information*, vol. 2, no. 1, p. 15023, Nov. 2016, arXiv: 1511.04206. [Online]. Available: http://arxiv.org/abs/1511.04206

[13] C. Murawski and P. Bossaerts, “How Humans Solve Complex Problems: The Case of the Knapsack Problem,” *Scientific Reports*, vol. 6, no. 1, pp. 1–10, Oct. 2016. [Online]. Available: https://www.nature.com/articles/srep34851

[14] R. Nowotniak and J. Kucharski, “GPU-based tuning of quantum-inspired genetic algorithm for a combinatorial optimization problem,” *Bulletin of the Polish Academy of Sciences, Technical Sciences*, vol. 60, pp. 323–330, 2012.

[15] ——, “Higher-Order Quantum-Inspired Genetic Algorithms,” 2014.

[16] X. Pan and T. Zhang, “Comparison and Analysis of Algorithms for the 0/1 Knapsack Problem,” *Journal of Physics: Conference Series*, vol. 1069, p. 012024, 2018.

[17] R. Rachwald, “Is banking online safer than banking on the corner?” *Computer Fraud & Security*, vol. 2008, pp. 11–12, 2008.

[18] T. Rakes, J. Deane, and L. Rees, “IT security planning under uncertainty for high-impact events,” *Omega*, vol. 40, pp. 79–88, 2012.

[19] M. Rocha and J. Neves, “Preventing Premature Convergence to Local Optima in Genetic Algorithms via Random Offspring Generation,” in *Multiple Approaches to Intelligent Systems*, 1999, pp. 127–136.

[20] A. Rohm and V. Swaminathan, “A Typology of Online Shoppers Based on Shopping Motivations,” *Journal of Business Research*, vol. 57, pp. 748–757, 2004.

[21] H. Wang, J. Liu, J. Zhi, and C. Fu, “The Improvement of Quantum Genetic Algorithm and Its Application on Function Optimization,” *Mathematical Problems in Engineering*, vol. 2013, p. 730749, May 2013. [Online]. Available: https://doi.org/10.1155/2013/730749