AJSE

American Journal of Science & Engineering

Volume 1 Issue 2

April 2020

American Journal of Science & Engineering (AJSE)
Society for Makers, Artists, Researchers and Technologists (SMART)
6408 Elizabeth Ave SE, Auburn 98092, Washington, USA
ISSN: 2687-9530 (Print) and 2687-9581 (Online)

Nonbinary Error-Detecting Hybrid Codes

Andrew Nemec
Andreas Klappenecker
nemeca@tamu.edu
klappi@cse.tamu.edu
Department of Computer Science & Engineering, Texas A&M University
College Station, Texas

***ABSTRACT*— Hybrid codes simultaneously encode both quantum and classical information, allowing for the transmission of both across a quan- tum channel. We construct a family of nonbinary error-detecting hybrid stabilizer codes that can detect one error while also encod- ing a single classical bit over the residue class rings Z_q inspired by constructions of nonbinary non-additive codes.**

CCS CONCEPTS

- ***Hardware→Quantum error correction and fault tolerance;***
- ***Theory of computation→Quantum information theory.***

KEYWORDS

quantum error-correction, nonbinary codes, hybrid codess.

1. INTRODUCTION

Hybrid codes allow for the simultaneous transmission of both quantum and classical information across a quantum channel. While it has long been known that simultaneous transmission can provide an advantage over the time-sharing of the channel for certain small error rates, see [4], most of the early work on the topic focused on information-theoretic results, see [7, 8, 25], while the problem of constructing finite-length hybrid codes remained largely overlooked.

The first examples of hybrid codes were given by Kremsky, Hsieh, and Brun [15], who introduced them as a generalization of entanglement-assisted stabilizer codes. Later, Grassl, Lu, and Zeng [5] gave multiple examples of small hybrid codes constructed using an approach inspired by the construction of nonadditive codeword stabilized quantum codes. Remarkably, these codes provide an advantage over optimal quantum codes regardless of the error rate.Recently, several families of hybrid codes have been constructed including several families constructed by the authors [20] for the Pauli channel using stabilzer pasting and a family constructed by Li, Lyles, and Poon [16] for fully correlated quantum channels.An operator-theoretic approach to hybrid codes has also been put forward in [2, 3, 18].

In [20], the authors constructed several families of binary hybridcodes with good parameters, including a family of [[n, n − 3:1, 2]]2error-detecting codes where n is odd. In this paper we provide ageneralization of this family to hybrid stabilizer codes over Zq,inspired by the non-additive nonbinary quantum codes constructedfrom qudit graph states by Hu et al. [9] and Looi et al. [17], as well as the family of single error-detecting codes given by Smolin, Smith, and Wehner [24].

1.1 Nonbinary Quantum Codes

A quantum code is a subspace of a Hilbert space that allows for the recovery of encoded quantum information even in the presence of arbitrary errors on a certain number of physical qudits. A quantum code has parameters n, K, d q if and only if it can encode a superposition of K orthogonal quantum states into the Hilbert space $Cq \otimes n \div Cqn$, while protecting the quantum information against all errors ocurring on less than d physical qubits.

Most generalizations of quantum codes from the binary alphabets to the case where $q > 2$ are constructed over the finite fields Fq , where q is a prime power, see [1, 10, 21]. In this paper, we instead follow [9, 17, 24] and construct codes over Zq for reasons that will become apparent in Section 2. Let $a, b \in Zq$. We define the unitary operators $X(a)$ and $Z(b)$ on Cq as

$$X(a)\,|x\rangle = |x + a\rangle \text{ and } Z(b)\,|x\rangle = \omega bx\,|x\rangle,$$

where ω = e2πi/q. The operators X(a) and Z(b)may be viewed as a generalization of the Pauli-X bit-flip error and the Pauli-Z phase error respectively. The set $\varepsilon=\{X(a)Z(b)|a,b\in\mathbb{Z}q\}$ forms a nice error basis on $\mathbb{C}^q$ see [11–13], meaning any error on a single qudit may be written as a linear combination of elements from ε. Additionally, any error on $\mathbb{C}q^n$may be written as a linear combination of errors from $\varepsilon_n=\varepsilon^{\otimes n}=\{E_1\otimes E_2\otimes \dots \otimes E_n|E_k\in\varepsilon,1\le k\le n\}$.By correcting errors from ε_n we are able to deal with arbitrary errors on the n qudits that are linear combinations of those errors. The weight wt(E) of an error Eε ε_n. is the number of non-identity tensor components it contains.

A quantum code C has the ability to detect an error $E \in \varepsilon_n$ if it either reports than an error occured or reports no error and returns a projection of the message back onto C. Formally, the Knill-Laflamme conditions tell us that an error E is detectable by a quantum code C if and only if $PEP = \lambda_E P$ for some scalar λ_E ,where P is the orthogonal projector onto C, see [14].

Stabilizer codes are perhaps the most important class of quantum codes, and are analogous to the linear and additive codes in classical coding theory (hence they are also refered to as additive codes). Stabilizer codes are completely determined by their stabilizer group S, an abelian subgroup of ε_n, and the code is defined as the subspace spanned by all joint eigenvectors of S with eigenvalue 1. Since this subspace will always have dimension $K = q^k$, we say the code has parameters $[\![n,k,d]\!]q$ to denote it as a stabilizer code.

1.2 Hybrid Codes

In addition to transmitting quantum information, we nowwant to simultaneously encode a classical message in with the encoded quantum state. A hybrid code has parameters$((n,K:M,d))_q$ if and only if it can simultaneously encode a superposition of K orthogonal quantum states as well as one of M different classical states into $(\mathbb{C}^q)^{\otimes n} \cong \mathbb{C}^{qn}$a Hilbert space of dimension qn, while protecting both the quantum and classical information against all errors of weight less than d.

An $((n,K:M,d))_q$ hybrid code C can be described by a collection of M orthogonal quantum codes Cm of dimension K, each indexed by a classical message $m \in [M] = \{0, 1, \dots, M - 1\}$.To transmit a quantum state φ and a classical messagem, we encode φ into the quantum code Cm. The Knill-Laflamme condiions for quantum codes can be generalized to hybrid codes, allowing us to characterize detectable errors: an error E is detectable by the hybrid code C if and only if

$$P_bEP_a=\begin{cases}-\lambda E, aPa & if\ a=b,\\ 0 & if\ a=b\end{cases} \quad (1)$$

for some scalar λE,a depending on both the error E and the classical message a, where Pa is the orthogonal projector onto the quantum code Ca. Equivalently, if $\{|c_i^{(m)}>\}$ are the codewords of the inner code Cm, we have that E is detectable by C if and only if

$$<c_j^{(b)}|E|c_i^{(a)}>=\boldsymbol{\lambda}_{E,a}\delta_{i,j}\delta_{a,b}. \quad (2)$$

If both the inner codes and outer code happen to be stabilizer codes, we say the code is a hybrid stabilizer code with parameters $\mathbb{I}n, k: m, d\mathbb{I}q$.In this case, the codes have some additional structure, so each inner code can be viewed as a translation from a seed code C0 by an operator tm ∈ En \ Z(S0), where Z(S0) is the centralizer in En of the stabilizer S0 of the seed code C0, so that Cm = tmC0. There are multiple simple constructions of hybrid codes using quantum codes described by Grassl et al. [5]:

1. Hybrid codes can be constructed using the following "trivial" constructions:

(1) Given an ((n,KM, d))q quantum code of composite dimension KM, there exisits a hybrid code with parameters ((n,K :M, d))q.

(2) Given an [[n, k :m, d]]q hybrid code with k > 0, there exists ahybrid code with parameters [[n, k − 1:m + 1, d]]q.

(3) Given an [[n1, k1, d]]q quantum code and an [n2,m2, d]q classical code, there exists a hybrid code with parameters [[n1 + n2, k1 :m2, d]]q.In each of these cases the sender is effectively substituting classical information for quantum information, which depending onthe context may be considered wasteful. In [5], Grassl et al. showedit was possible to construct genuine hybrid codes that provide an advantage over these simple codes, and provided examples of suchcodes found using an exhaustive search of small parameters. In [20]the authors constructed several infinite families of genuine hybridcodes, including a family of binary single error-detecting codeswhich we generalize to the nonbinary case in the next section.

2 FAMILY OF HYBRID CODES OVER Zq

The first good non-additive quantum code (that is a quantum code that is not a stabilizer code) was the ((5, 6, 2))2 code given by Rainset al. [23]. This code outperforms the optimal [[5, 2, 2]]2 stabilizer code, and was further generalized by Rains [22] into a family of odd-length non-additive codes that outperform optimal stabilizer codes. However, for an odd-length ((n,K, 2)) quantum code we have the following bound:

$$K \le 2^{n-2}\left(1-\frac{1}{n-1}\right) \quad (3)$$

and many families of codes that approach this bound have been constructed. In [20], the authors gave a construction for a familiy of hybrid stabilizer codes with parameters [[n, n − 3:1, 2]]2 that beat this bound.

Nonbinary quantum codes with similar parameters were hinted at by Rains in [22], and first given by Smolin et al. [24] as a generalization of their family of non-additive binary codes. Soon after,further families were constructed by Hu et al. [9] and Looi et al. [17]using qudit graph states. All of these families are codes over integerrings rather than finite fields, and our construction of nonbinary hybrid stabilizer codes will follow in their footsteps. The reasonwe choose to construct codes over Zq rather than Fq is due to the following result of Grassl and Rötteler:

Theorem 1 ([6, Theorem 12]). Let q > 1 be an arbitrary integer, not necessarily a prime power. Quantum MDS codes C = [[n, n − 2, 2]]q exist for all even length n, and for all length n ≥ 2 when the dimension q of the quantum systems is an odd integer or is divisible by 4.

While the construction below will certainly produce a hybrid stabilizer code when q . 2 mod 4, it will not be a genuine hybrid code, as the previous theorem implies that there will be an [[n, n − 2, 2]]q stabilizer code that can be transformed into a hybridcode using the first construction in Proposition 1. When q = 2,Equation 3 tells us that there can be no [[n, n − 2, 2]]2 quantum code, implying that the family given in [20] is indeed genuine. To the best of our knowledge there are no known [[n, n − 2, 2]]q codes when q = 4r + 2, which is why the codes using the construction below may in fact be genuine. However, since F4r+2 does not exist except when r = 0, we instead construct our codes over Zq.

Proposition 2. Letn be odd. Then there exists an [[n, n − 3:1, 2]]Zq hybrid code.

Proof. Let a, b ∈ Znq , m ∈ Zq, and ω a primitive q-th root of unity. Define the following states:

$$|\emptyset a,b>=\frac{1}{q^n}\sum_{c\in\mathbb{Z}_q^{2n}}\omega^{\sum_{i=1}^{n}(c_{2i-1}-a_i)}|C>$$

$$|\Psi_m>=\frac{1}{\sqrt{q}}\sum_{c\in\mathbb{Z}_q}\omega^{mc}|c>$$

Define the inner code Cm as follows:

$$C_m=\langle|\phi_{a,b}\rangle\otimes|\psi_m\rangle|a,b\in\mathbb{Z}_q\ \textstyle\sum_{i=1}^{n}a_1=0,\sum_{i=1}^{n}b_i=m\rangle$$

The state $|\Psi_{a,b}\rangle$ is the tensor product of two-qubit states of the form

$$|\phi_{a_i,b_i}\rangle=\frac{1}{q}\sum_{c\epsilon\mathbb{Z}_q^2}\omega^{(c_1-a_i)(c_2-b_i}|c\rangle$$

For two of these states $|\phi_{a_i,b_i}\rangle, |\phi_{a_i',b_i'}\rangle$ we have

$$\langle\phi_{a_{i,b_i}}|\phi_{a_{i,}'\ b_{i,}'}\rangle=\frac{1}{q^2}\sum_{c\in\mathbb{Z}_q^2}\omega^{(c_1-a_i')(c_2-b_i')-(c_1-a_i)(c_2-b_i)}$$

$$=\frac{\omega^{a_i'b_i'-a_ib_i}}{q^2}\sum_{c\in\mathbb{Z}_q^2}\omega^{c_1(b_i'-b_i)+c_2(a_i'-a_i)}$$

$$=\frac{\omega^{a_i'b_i'-a_ib_i}}{q^2}\left(\sum_{c_1\in\mathbb{Z}_q}\omega^{c_1(b_i'-b_i)}\right)\left(\sum_{c_2\in\mathbb{Z}_q}\omega^{c_2(a_i'-a_i)}\right)$$

$$=\begin{cases}1\ if a_i=a_i' and b_i=b_i'\\ 0 \qquad \text{otherwise}\end{cases}$$

Therefore for the full states $|\phi a,b\rangle$, $|\phi a',b'\rangle$ we have the same:

$$\langle\phi_{a,b}|\phi_{a',b'}\rangle=\begin{cases}1\ if a=a' and b=b'\\ 0 \qquad \text{otherwise}\end{cases}$$

Similarly, for $|\psi_m\rangle$, $|\psi_{m'}\rangle$ we have

$$\langle\psi_m|\psi_{m'}\rangle=\begin{cases}1 & if\ m=m'\\ 0 & \text{otherwise}\end{cases}$$

Thus all of the codewords are orthogonal to one another.

Consider two codewords $|\phi_{a,b}\rangle\otimes|\psi_m\rangle, \phi_{a',b'}\rangle\otimes|\Psi_{m'}\rangle$Suppose

that a Pauli-X(u) error occurs on the first n −1 qudits. Without loss of generality, we can assume that the error occurred on either the first or second qudit.If $m \neq m', a_i \neq a_i', or\ b_i \neq b_i'\ for\ 1 < i \leq n$, then

$$(\langle\psi_{a,b}|\otimes\langle\psi_m|)X(u)(|\phi_{a',b'}\rangle\otimes|\psi_{m'}\rangle) = 0$$

by the orthogonality relations above. Therefore we can restrict our attention to the case where $m = m', a_i = a_i'\ and\ b_i = b_i'$
for (1< i≤ n). We note that these restrictions along with the requirement that the $a_i$and a_i'sum to 0 and b_i and b_i'sum to m and m' respectively completely determine the values $a_1$and$b_1$and in particular we must have $a_1 = a_1'$and $b_1 = b_1'$. If the error
occurred on the first qudit, we have

$$\langle\phi_{a_1,b_1}|X(u)|\phi_{a_1,b_1}\rangle$$

$$= \frac{1}{q^2}\left(\sum_{c\in\mathbb{Z}_q^2}\omega^{-(c_1-a_1)(c_2-b_1)}\langle c_1 c_2|\right)\left(\sum_{c\in\mathbb{Z}_q^2}\omega^{(c_1-a_1)(c_2-b_1)}|(c_1 + u)c_2\rangle\right)$$

$$= \frac{1}{q^2}\left(\sum_{c\in\mathbb{Z}_q^2}\omega^{-(c_1-a_1)(c_2-b_1)}\langle c_1 c_2|\right)\left(\sum_{c\in\mathbb{Z}_q^2}\omega^{(c_1-a_1-u)(c_2-b_1)}|c_1 c_2\rangle\right)$$

$$= \frac{1}{q^2}\sum_{c_2\epsilon\mathbb{Z}_q^2}\omega^{u(b_1-c_2)}$$
$$= \begin{cases} 1 & if\ \ u = 0 \\ 0 & \text{otherwise} \end{cases}$$

A similar argument holds if the error occurs on the second qudit,thus the code can detect any single Pauli-X(u) error that occurs on the first n − 1 qudits. Now suppose that a Pauli-Z(v) error occurs on the first n − 1 qudits. As above, we restrict our attention to the case where $a = a', b = b'and\ m = m'$,and the error occurs on one of the first two qudits.If the error occurs on the first qudit we have

$$\langle\phi_{a_{1,b_1}}|Z(v)|\phi_{a_1,b_1}\rangle$$
$$= \frac{1}{q^2}\sum_{c\in\mathbb{Z}_q^2}\omega^{(c_1-a_1)(c_2-b_1)-(c_1-a_1)(c_2-b_1)+vc_1}$$
$$= \frac{1}{q}\sum_{c\in\mathbb{Z}_q^2}\omega^{vc_1}$$
$$= \begin{cases} 1 & if\ \ v = 0 \\ 0 & \text{otherwise} \end{cases}$$

The same argument holds if the error occurs on the second qudit,thus the code can detect any single Pauli-Z(v) error that occurs on the first n − 1 qudits. Now suppose that a Pauli error E occurs on the last qudit. If $a \neq a', b \neq b', or\ m \neq m'$ then the orthogonality of the first n − 1 qudits gives us

$$(\langle\phi_{a,b}|\otimes\langle\psi_m|)E(|\phi_{a',b'}\rangle\otimes|\Psi_m'>) = 0$$

so againwe only need to examine the case where the two codewords are the same.

If we have a Pauli-X(u) error on the last qudit we have

$$\langle\psi_m|X(u)|\psi_m\rangle = \frac{1}{q}\left(\sum_{c\in z_q}\omega^{-mc\langle c|}\right)\left(\sum_{c\in z_q}\omega^{mc|c+u\rangle}\right)$$
$$= \frac{1}{q}\sum_{c\in\mathbb{Z}_q}\omega^{-mu}$$
$$= \omega^{-mu}$$

meaning that the error is degenerate. Note that since the value depends on the classical informationm, each inner code can detect the error but the outer code (as a quantum code) cannot.
If a Pauli-Z(v) error occurs on the last qudit we have

$$\langle\psi_m|Z(v)|\psi_m\rangle = \frac{1}{q}\left(\sum_{c\in z_q}\omega^{-mc\langle c|}\right)\left(\sum_{c\in z_q}\omega^{mc+vc}\ \ |c\rangle\right)$$
$$= \frac{1}{q}\sum_{c\in\mathbb{Z}_q}\omega^{vc}$$
$$= \begin{cases} 1 & if\ \ v = 0 \\ 0 & \text{otherwise} \end{cases}$$

We also mention in passing that this construction can be generalized further to codes over Frobenius rings by replacing the primitive root of unity by an irreducible additive character of the additive group of the ring [19].

3 CONCLUSION AND DISCUSSION

Hybrid codes simultaneously transmit both quantum and classical information across quantum channels, and can provide an advantage over using quantum codes for simultaneous transmission. We have generalized a family of single error-detecting codes constructed in [20] from the binary case to the nonbinary case. While it is known that the construction gives genuine hybrid codes when q = 2, the existence of quantum codes with the similar parameters when q ≡ 0, 1, 3 mod 4 means the construction does not produce genuine hybrid codes in all cases. One open question is whether or not the codes given by the construction are always genuine when q ≡ 2 mod 4. As the code family here is the only construction of nonbinary hybrid codes, further investigation is needed.

REFERENCES

[1] Alexei Ashikhmin and Emanuel Knill. 2001. Nonbinary Quantum Stabilizer Codes. IEEE Trans. Inform. Theory 47, 7 (2001), 3065–3072.
[2] Cédric Bény, Achim Kempf, and DavidW. Kribs. 2007. Generalization of Quantum Error Correction via the Heisenberg Picture. Phys. Rev. Lett. 98, 10 (2007), 100502.
[3] Cédric Bény, Achim Kempf, and David W. Kribs. 2007. Quantum error correction of observables. Phys. Rev. A 76, 4 (2007), 042303.
[4] I. Devetak and P. W. Shor. 2005. The Capacity of aQuantum Channel for Simultaneous Transmission of Classical and Quantum Information. Commun. Math. Phys. 256, 2 (2005), 287–202.
[5] Markus Grassl, Sirui Lu, and Bei Zeng. 2017. Codes for Simultaneous Transmission of Quantum and Classical Information. In Proc. 2017 IEEE Int. Symp. Inform.Theory (ISIT). Aachen, Germany, 1718–1722.
[6] Markus Grassl and Martin Rötteler. 2015. Quantum MDS Codes over Small Fields.In Proc. 2015 IEEE Int. Symp. Inform. Theory (ISIT). Hong Kong, China, 1104–1108.

[7] Min-Hsiu Hsieh and Mark M. Wilde. 2010. Entanglement-Assisted Communication of Classical and Quantum Information. IEEE Trans. Inform. Theory 56, 9 (2010), 4682–4704
[8] Min-Hsiu Hsieh and Mark M. Wilde. 2010. Trading Classical Communication,Quantum Communication, and Entanglement in Quantum Shannon Theory. IEEE Trans. Inform. Theory 56, 9 (2010), 4705–4730.
[9] Dan Hu,Weidong Tang, Meisheng Zhao, Qing Chen, Sixia Yu, and C. H. Oh. 2008.Graphical nonbinary quantum error-correcting codes. Phys. Rev. A 78, 1 (2008),012306.
[10] Avanti Ketkar, Andreas Klappenecker, Santosh Kumar, and Pradeep Kiran Sarvepalli. 2006. Nonbinary Stabilizer Codes Over Finite Fields. IEEE Trans.Inform. Theory 52, 11 (2006), 4892–4914.
[11] Andreas Klappenecker and Martin Rötteler. 2002. Beyond Stabilizer Codes I: Nice Error Bases. IEEE Trans. Inform. Theory 48, 8 (2002), 2392–2395.
[12] Andreas Klappenecker and Martin Rötteler. 2003. Unitary Error Bases: Constructions,Equivalence, and Applications. In Applied Algebra, Algebraic Algorithms,and Error Correcting Codes - Proceedings 15th International Symposium, AAECC-15,Toulouse, France, Marc Fossorier, Tom Høholdt, and Alain Poli (Eds.). Springer-Verlag, 139–149.
[13] E. Knill. 1996. Non-binary Unitary Error Bases and Quantum Codes. (Jun. 1996).Los Alamos National Laboratory Report LAUR-96-2717.
[14] Emanuel Knill and Raymond Laflamme. 1997. Theory of quantum errorcorrecting codes. Phys. Rev. A 55, 2 (1997), 900–911.
[15] Isaac Kremsky, Min-Hsiu Hsieh, and Todd A. Brun. 2008. Classical enhancement of quantum-error-correcting codes. Phys. Rev. A 78, 1 (2008), 012341.
[16] Chi-Kwong Li, Seth Lyles, and Yiu-Tung Poon. 2019. Error correction schemes for fully correlated quantum channels protecting both quantum and classical information. (Jun. 2019). arXiv:1905.10228v2 [quant-ph].
[17] Shiang Yong Looi, Li Yu, Vlad Gheorghiu, and Robert B. Griffiths. 2008. Quantumerror-correcting codes using qudit graph states. Phys. Rev. A 78, 4 (2008), 042303.
[18] Shayan Majidy. 2018. A unification of the coding theory and OAQEC perspective on hybrid codes. (Jun. 2018). arXiv:1806.03702 [quant-ph].
[19] Sushma Nadella and Andreas Klappenecker. 2012. Stabilizer Codes over Frobenius Rings. In Proc. 2012 IEEE Int. Symp. Inform. Theory (ISIT). Cambridge, Massachusetts,165–169.
[20] Andrew Nemec and Andreas Klappenecker. 2019. Infinite Families of Quantum-Classical Hybrid Codes. (Nov. 2019). arXiv:1911.12260 [quant-ph].
[21] Eric M. Rains. 1999. Nonbinary Quantum Codes. IEEE Trans. Inform. Theory 45,6 (1999), 1827–1832.
[22] Eric M. Rains. 1999. Quantum Codes of Minimum Distance Two. IEEE Trans.Inform. Theory 45, 1 (1999), 266–271.
[23] Eric M. Rains, R. H. Hardin, Peter W. Shor, and N. J. A. Sloane. 1997. Nonadditive Quantum Code. Phys. Rev. Lett. 79, 5 (1997), 953–954.
[24] John A. Smolin, Graeme Smith, and Stephanie Wehner. 2007. Simple Family of Nonadditive Quantum Codes. Phys. Rev. Lett. 99, 13 (2007), 130505.
[25] Jon Yard. 2005. Simultaneous classical-quantum capacities of quantum multiple access channels. Ph.D. Dissertation. Stanford University, Stanford, CA. arXiv

Meeting the Problems of Traffic Congestion in Beirut Southern Entrance in Lebanon

Safaa Ahmad Hajjoul
Department of Indunstrial Engineering
Beirut Arab University Lebanon ,
Debbieh safaahajjoul@hotmail.com

Dr. Ali Ghandour
Researcher
National Council for Scientific Research Lebanon,
Mansorieh gandour.ali@gmail.com

Ranime El Hadi
Department of Indunstrial Engineering
Beirut Arab University Lebanon Debbieh
ranime.elhady@gmail.com

***Abstract*— In this paper, we tried to model the natural behavior of traffic in khalda, which is the most important area in Beirut Southern Entrance (BSE). We used queuing models to model the queuing system and to calculate the average waiting time in queue in addition to computing cost metrics and this was to quantify traffic congestion.**

Keywords*— *Traffic Congestion- Queuing models- Cost Analysis

1. INTRODUCTION

Every morning, Beirut capital welcomes commuters from all regions with heavy traffic on its entrances. The congregation of governmental divisions, the majority of country's company offices and workplaces, educational institutions, entertainment centers, all in the capital has resulted of a large demand on commuting into the city. However, with lack of sustainable and efficient transport system.

1.1 Definition

Traffic congestion is one of the biggest problems that most cities are facing. It is a common problem among developed and undeveloped countries so it is a global problem. There is NO universally accepted definition of traffic congestion because it is both a physical and a relative phenomenon.

As a Physical phenomenon traffic congestion is described as a situation where demand for road space exceeds supply while as a Relative phenomenon, it is described as the difference

Lebanese citizens are relying heavily on private vehicles for transportation. On other hand, the city is undergoing population increase and with poor roads infrastructure all factors combines to create an intolerable congestion that has become gradually a part of the citizens' quotidian life between road performance and road user's expectations.

1.2 Study Area

Beirut has three main entrances:

The Northern Entrance which has 300, 000 commuter/day, the Southern Entrance which has 100, 000 commuter/day, and the Eastern Entrance which has 70, 000 commuter/day but actually there is about 500,000 cars in wandering in Beirut itself so the Total number of cars in Beirut is 1, 000, 000 car/day. In this study, I focused on Beirut southern entrance. The southern access roads are Khalda, Oozai, Cola, Chatila, Sports City and Cocody. The study took place in Khalda.

2. Methodology

The key adopted methodology is summarized as follows:

a. Identifying the sources of Traffic congestion in Beirut southern entrance.
b. Focus on two main sources
c. Pre-screening the study area using high-resolution satellite images.
d. Site survey
e. Collect data and record videos in the peak hours
f. Simulate the problems using Arena software
g. Devise appropriate Analytical Model
h. Compute Cost Metrics
i. Suggest potential solutions

2.1 Sources of traffic congestion in Beirut Southern Entrance (BSE)

I classified the sources of traffic congestion in Beirut Southern entrance into three categories:

A. Sources in the Eastern route(E)
B. Sources in the Western route (W)
C. Common sources (C)

A. Sources of traffic congestion in Beirut Southern entrance in the Eastern route:

E1: Improper design of highway entrances and exits
E2: Presence of bus stations on the highway
E3: Using obstacles that divide the highways into branches as a picking and landing stations.

B. Sources of traffic congestion in Beirut southern entrance in the Western route are:

W1: Presence of peddlers on the highway
W2: Poor traffic management and traffic lights
W3: Presence of the Bus stations on the highway
W4: Presence of commercial buildings on the highway.

C. Common sources in the eastern and western route:

C1: Existence of water filters and poor road infrastructure
C2: Multi-lanes

.

What problems to focus on?

The main sources that we will focus on are E1, and W4.

E1: E1 is an exit in khalda near Rammal's super market where vehicles are using this exit as an entrance to the highway, which in turn reduces the capacity of the highway.

The readings were from 7:00 Am until 9:00 Am.

W4: presence of commercial buildings (Harkous Chicken) on the highway is illegal, and unaccepted, and it is the main source of traffic congestion in the Western route since it open a direct access to the highway. The readings were from 4:00 Pm until 6:00 Pm.

2.2 Data collection Plan

For E1, the collected data was:

1- Number of cars arriving to the exit per unit time on L3 (arrival rate)
2- Number of cars entering the exit per unit time from L3
3- Percentage of cars entering the exit (%)
4- Number of cars getting out from the exit
5- Time needed to reach the top of the exit
6- Merging time when no merging is occurring (μ0)
7- Merging time when merging is occurring (μB))

For W4, the collected data was:

1- Number of cars arriving to the entrance of the parking per unit time on L3 (arrival rate)
2- Number of cars entering P1 (inside the parking)
3- Number of cars parking in P2 (outside the parking)
4- Time spent in P1 and P2
5- Time needed to park
6- Time spent in the restaurant
7- Time needed to leave the parking
8- Merging time when no merging occurs' (μ0)
9- Merging time when merging is occurring (μB))

After collecting data, we entered it into Simulation Arena Software to model the problems and understand them better.

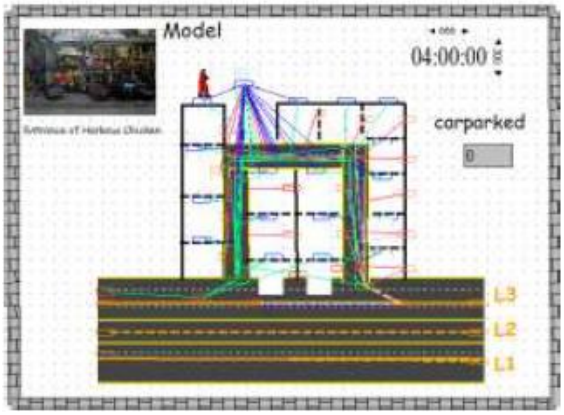

Figure 1: The model built in simulation Arena for W4

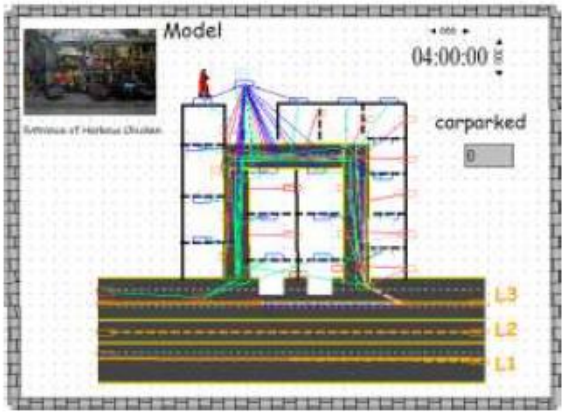

Figure 2: The model built in Simulation Arena for E1

2.2 Analytical model

Queuing theory is the mathematical study of waiting lines and it used, as Analytical model .The purpose is to calculate the **average waiting time**. Queuing models are often identified by three values: arrival rate assumption, the departure rate assumption, and the number of departure channels.

After reading more about queuing models, the Suitable model is **M/G/1**, which describes the vehicular merging where M stands for Markovian and G for general with one departure channel

The system will be based on the following assumptions:

- Constant traffic flow for a short time period **(λ)**
- The time between servicing two cars is of exponential distribution
- The driver has to wait for the other drivers that are in front of the line to be served first (FIFO)

Formulas of M/G/1:

a. Arrival rate (λ) = $K * Ø/T$

Where:

- T is the number of tollbooths
- K: number of Linear streams (lanes) after the vehicles leave the tollbooths.
- Ø: traffic flow rate which is the number of cars passing a single point per time unit (per one lane)

b. Expected number of drivers in the system

$$L(\lambda) = \sum_{0}^{\infty} iP_i = \frac{\lambda}{\mu_B - \lambda} + \frac{\lambda(\mu_B - \mu_0)}{\lambda(\mu_B - \mu_0) + \mu_0\mu_B}$$

c. Average waiting time in the system

$$t_{sys}(\lambda) = \frac{L(\lambda)}{\lambda} = \frac{1}{\mu_B - \lambda} + \frac{\mu_B - \mu_0}{\lambda(\mu_B - \mu_0) + \mu_0\mu_B}$$

d. Average wasted time of a driver at a merging point

$$t_{diff}(\lambda) = t_{sys}(\lambda) - \frac{1}{\mu_0} = \frac{1}{\mu_B - \lambda} + \frac{\mu_B - \mu_0}{\lambda(\mu_B - \mu_0) + \mu_0\mu_B} - \frac{1}{\mu_0}$$

2.3.1 Analytical model for W4:

In w4, we have three incoming lanes (**L1, L2, and L3)** and a side clearance from the right side. Vehicles that wants to enter "Harkous Chicken" are usually on the right lane which is named in our model as **L3**.When vehicles arrives to the entrance of the parking of Harkous Chicken , they will first check if there is any empty spot inside the parking . In case there is an empty spot, then they will park inside **(P1)** and if there isn't any empty spot then they will park outside the parking **(P2).** when we say outside , this means that they will park in the space that this between the side clearance and **L3** which in turns will decrease the capacity of **L3** and will oblige the incoming vehicles to merge with **L2** taking into consideration that the vehicles cannot merge unless there are no cars coming from **L2**. In **W4**, the arrival rate before the occurrence of merging is to be assumed as uniformly distributed.

The two incoming lanes are treated as one queue, and the total merging process In **W4** is considered as 2-to-1 merging points. The merging point on **L2** is receiving a traffic stream coming from **L3** with total flow of **ø**. Traffic flow is usually considered roughly constant in the highway.

ø = 2000 veh/hr. = 13.33 veh/min
K=1 and T=2
λ = k*ø/T= 1000 veh/hr. = 16.67 veh/min
μ_B=3.22 sec
Service rate = 18.64 veh/min = 1118.24 veh/hr
μ_0=1.44 sec

Service rate =41.72 veh/min =2502.81 veh/hour

By substituting:

L (λ) **=8 drivers**

$tsys(\lambda)$=**0.45 min/veh**

$tdiff(\lambda)$=**0.5 min/veh**

2.3.2 Analytical model for E1

Here we have three incoming lanes (**L1, L2 and L3)** and a side clearance and as we observed the side clearance is used as a forth lane in case of traffic congestion. The exit that is near the "Rammal's" is used as an entrance to the highway, and here there are two lines coming from this exit, the first line takes **L3** directly which in turns obliges the incoming vehicles on **L3** to stop when merging is occurring. while vehicles taking the second line takes a lane and a half while merging which is considered as a blockage area for the incoming vehicles on **L2 and L3** which leads to lane changing and merging for **L2** and **L1.** Note that the flow of vehicles entering or exiting at any given interchange is usually much less than the flow on the mainline roadway

The first merging point (M1): which the merging point between the exit and L3.

$ø_1$ = 28 veh / min = 1680 veh / hr.

$ø_2$= 24 veh/min =1440 veh/hr.

Total Ø= 52 veh/min=3120 veh/hr.

K=1, T=2

λ= k*ø/T = 26 veh/min=1560 veh/hr.

μ_B: = 7.46 veh/min

μ0: 11.56 veh/min

By substituting:

L (λ) =4 drivers **= 4 drivers**

$tsys(\lambda)$ **=0.15 min/veh**

$tdiff$ (λ) **=0.15 min/veh**

The second merging point **(M2):** This is the merging point between L2 and L3.

ø = 28 veh/ min =1680 veh/hr.

K=1, T=2

λ=k*ø/T=14 veh/min =840 veh/hr.

μB = 19.08veh/min

μ0 = 27.28 veh/min

By substituting:

L(λ) **= 3 drivers**

$tsys(\lambda)$ **=0.18 min/veh**

$\boldsymbol{tdiff(\lambda)}$**= 0.13 min/veh**

2.4 Cost analysis

This section provides the calculation of estimated direct economic costs of traffic congestion in Beirut Southern entrance using analytical models in E1 and W4.The most important parameter is the average waiting time, which was calculated using the formulas of M/G/1 queuing model

2.4.1 Cost Analysis for W4

.

1- Average waiting lost time

$tsys(\lambda)$= 0.45 min / veh

Average loss time rate **=450 minutes/hr.**

If we convert this time into cost according to the minimum wages in Lebanon, we will find that this lost time has a cost and can effect thoroughly on the person's work.

- Time Lost near Harkous chicken is 450 min /hr.
- If we are taking the minimum Lebanese wages which is 900,000 L.L and 5 business days

=900,000 $L.L25\ days*8\ hrs*60\ min/$

=**75 L.L /min=4500 L/L / hour**

The cost of average time lost is nearly (assuming one person/car): 75

L.L/min *450 min **=33,750 L.L /hr.**

2- Fuel consumption cost

Vehicle Type	Idling Fuel Use (gal/h)
Passenger car (ford focus)	0.16
Passenger Car (Volkswagen Jetta)	0.17
Passenger Car (Ford Crown Victoria)	0.39
Medium Heavy Truck	0.84
Delivery Truck	0.84
Tow Truck	0.59
Medium Heavy Truck	0.44
Transit Bus	0.97
Combination Truck	0.49
Bucket Truck	0.9F
Tractor-Semitrailer	0.64

Table 1: The idling fuel use (gal/h)

Average idling fuel use = 0.59 gallons / hour=0.0097 gallons/minute= 0.036879088 Liters/ min

Fuel	**fuel price/20 liter L.L**		
98-Octane	22,900		
95-Octane	22,30		
Diesel	13,500		
Average fuel cost/20 Liters	19,567		
Average fuel cost/1 Liter	978.3333333		
total time lost / hour (min)	**fuel consumed in 1 minute (L/min)**	**Total fuel consumed /hr. (L)**	**Total fuel cost in L.L**
450	0.036879088	16.5955896	16236.01849

Table 2: cost of total fuel consumed in queue (W4)

1 gallon=3.7854118 Liters

Total fuel cost /hour =987.33333 L.L*16.5955896 liters/ hour = **16,236.01849 L.L** /hr

3-Airpollution

Based on studies:

Pollutant/Fuel	**Emission (gram)**	**cost $/ton**	**Emission /ton**
VOC	1.034	$2,392	0.000001034
CO2	368.4	$15	0.0003684
NOX	0.693	$10,293	0.000000693
PM10	0.0044	$10,868	4.4E-09
PM2.5	0.0041	$63,339	4.1E-09
Gasoline consumption (gallons)	0.04149		

The Combustion of **0.1570566 Liters (0.04149 gallons)** of Gasoline emits

Table 3: Average Emissions and Fuel Consumption for Passenger Cars

Moreover, if 1 gallon has a capacity of 3.78541 liters, then the gasoline consumption equals 0.15705661 Liters.

1ton=1,000,000 grams

Pollutant/Fuel	Emissions/ton	cost/ton
Voc	0.000109259	$0.26
Co2	0.038927449	$0.58
NOx	7.32267E-05	$0.75
PM10	4.64932E-07	$0.01
PM2.5	4.33232E-07	$0.03
total costs $	$1.63	total costs $
total costs L.L	2447.212	total costs L.L

Table 4: The calculation of emissions/ton and their cost for Voc, co2, NOx, Pm10, and Pm2.5 in W4

- The total estimated costs of emitted gases of 16.5955896 Liters/hour is = 1.63 $
- The cost of emitted gases/hour = 1.63 * 1500 =**2,445 L.L / hr.(assuming one person in the car)**

So then, the total cost (assuming one commuter \car)
=Average time lost + fuel cost +gas emission cost
=33,750 L.L/hr. + 16,236.01849 L.L/hr. + 2,445 L.L/hr.
=52,431.01849 L.L/hr.
The actual cost is much higher!

2.4.2 Cost analysis for E1

For E1, the cost analysis will be for M1 and M2.

1- Average waiting lost time

$t_{sys}(\lambda)$ E1, M1 = **0.15 min / veh**
Average loss time rate =**234 minutes/hr.**
$t_{sys}(\lambda)$ E1, M2 = **0.18 min / veh**
Average loss time rate =**151.2 minutes/hr**.
Total time lost in E1=6.42 min/min= **385.2 min/hr**.

According to the minimum wages in Lebanon:

The cost of average time lost in M1 is nearly: 75 L.L/min *234 min =**17,550 L.L /hr.**

The cost of average time lost in M1is nearly: 75 L.L/min *151.2 min =**11,340 L.L /hr.**

The total cost of average waiting time lost in E1 (assuming one person in the car) = 28,890 LL/hr.

2- Fuel Consumption cost

	Fuel	fuel price/20 liter L.L		
	98-Octane	22,900		
	95-Octane	22,300		
	Diesel	13,500		
	Average fuel cost/20 Liters	19,567		
	Average fuel cost/1 Liter	978.3333333		
	total time lost / hour (min)	fuel consumed in 1 minute (L/min)	Total fuel consumed /hr. (L)	Total fuel cost in L.L
for M1	234	0.036879088	8.629706592	8442.7296
for M2	151.2	0.036879088	5.576118106	5455.3022

Table 5: The cost of total fuel consumed in queue E1 (M1 and M2)

The total fuel cost in M1 and M2 =**13, 898.0318 L.L/h**

3-Air pollution

Pollutant/Fuel	Emissions/ton	cost/ton
Voc	**9.35256E-05**	**$0.22**
Co2	**0.033321897**	**$0.50**
NOx	**6.26821E-05**	**$0.65**
PM10	**3.97981E-07**	**$0.00**
PM2.5	**3.70846E-07**	**$0.02**
total costs $		**$1.40**
total costs L.L		**2094.814**

Table 6: Calculation of emissions/ton and their cost for Voc, co2, Nox, Pm10, and Pm2.5 in E1 (M

- The total estimated costs of emitted gases of 14.205824698 Liters /hour is = 1.40 $
- The cost of emitted gases /hr. = **2100 L.L/hr.(on our health assuming one person in every car)**

So then the total cost (assuming 1 commuter/car) =average time lost + fuel cost +gas emission cost
=28, 890 L.L/hr. + 13, 898.0318 L.L/hr. + 2100 L.L/hr.
=44,888.0318 L.L/hr.
The Actual cost is much higher!

5.4 Suggested Solutions

As a conclusion, the problem of traffic will going to be worst if the existing state will continue in this ways especially after the Syrian asylum in Lebanon and after doing many researches about this problem preventing traffic congestion is an impossible solution in a growing country such as Lebanon. All what we can do is to eliminate traffic congestion and there are various ways.

With respect to W4, the commercial buildings entry should be from the old marine road and exiting cars should be forbidden from merging with the highway and instead directed to the old marine old. The government should play an additional role, which is to issue a decision, which is any car that uses the side clearance, as a parking will be punished by paying financial penalty, and to forbid new commercial buildings from building closely to the highway.

With respect to E1, the suitable solution is to apply the rules by forbidding using this exit as an entrance to the highway.

Some other solutions is the decentralization of the main government divisions in Beirut and we may change the shift of theschools, which will decrease the traffic volume.

Acknowledgement:

Amudapuram Mohan Rao, Kalaga Ramachandra Rao, 2012.International Journal for Traffic and Transport Engineering, MEASURING URBAN TRAFFIC CONGESTION

Transport Engineering, MEASURING URBAN TRAFFIC CONGESTION

REFERENCES:

Anthony Downs. Still Stuck in Traffic. (Washington D.C.: Brookings Institution) 2004.

Beirut southern entrance phases 2002

Beirut, The Pearl of the Middle East Archived 4 October 2013 at the Way back Machine.

Beylich (1978), Elements of a Kinetic Theory of Traffic Flow. Proceedings of the Eleventh International Symposium on Rarefied Gas-Dynamics, Vol 1, Cannes, France, 129-138.

Botma, H. (1978), State-of-the-Art report "Traffic Flow Models" (in Dutch). Research Report R-78-40, SWOV

Ceballos, Gustavo and Owen Curtis, \Queue Analysis at Toll and Parking Exit Plazas: A Comparison Between Multi-server Queuing Models and Traffic Simulation".http://www.ptvamerica.com/docs/VISSIMQueueAnalysis.pdf

Delaware Department of Transportation, \Dedication Ceremony Held for Opening of E-ZPass Express Lanes at the Dover Toll Plaza". http://www.deldot.net/public. Ejs? Command=PublicNewsDisplay&id=1823&month=5&year=2004.

Francois, M.I., & Willis, A., (1995). Developing Effective Congestion Management Systems. Federal Highway Administration, Technical Report No.8, p.22

Gross, Donald and Carl M. Harris, Fundamentals of Queueing Theory, Wiley, New York, 1998.

Hock, Ng Chee, Queueing Modeling Fundamentals, Wiley, New York, 1996.

Hoogendoorn, S.P., and P.H.L. Bovy. (2000b). Modelling Multiple User-Class Traffic Flow. Transportation Research B (34)2, 123-146.

Kerner, B.S., Konhäuser, P., and Schilke, M. (1996), A new approach to problems of traffic flow theory. In: Lesort, J.B. (ed), Proceedings of the 13th International Symposium of Transportation and Traffic Theory, Lyon, 119-145.

Khan, ABM S.; Clark, N.N.; Gautam, M.; et al. (2009). "Idle Emissions from Medium Heavy Duty Diesel and Gasoline Trucks." Journal of the Air & Waste Management Association (59:3) 354–359.

Klar, A., and Wegener (1998), A Hierarchy of Models for Multilane Vehicular Traffic I & II: Modelling. SIAM Journal of Applied Mathematics.

[Kühne, R.D. (1991). Traffic Patterns in unstable Traffic Flow on Freeways. Highway Capacity and Level of Service. Brannolte (ed.), Rotterdam.

Lebanon State of the Environment Report, Ministry of Environment/LEDO, chapter 5: Transport.

Leutzbach, W. (1988) .An introduction to the theory of traffic flow, Springer-Verlag, Berlin.

Lighthill, M.H., and G.B. Whitham (1955). On kinematic waves II: a theory of traffic flow on long, crowded roads. Proceedings of the Royal Society of London series A, **229**, 317-345

Lomax,T., Turner, S., Shunk, G., Levinson, H.S., Pratt, R.H., Bay, P.N., & Douglas, G.B. (1997). Quantifying Congestion. Final Report. National Cooperative Highway Research Program, Transportation Research Board, p.184

Mahmud, K., Gope, K., & Chowdhury, S. M. R. (2012). "Possible causes and solutions of traffic jam and their impact on the economy of Dhaka City". Journal of Management and Sustainability, 2(2).

May, A D (1990). Fundamentals of traffic flow and queuing theory, chapter5.

Munjal, P., and J. Pahl (1969). An Analysis of the Boltzmann-type Statistical Models for Multi-Lane Traffic Flow. Transportation Research **3**, 151-163.

Nagel, K. (1996). Particle Hopping Models and Traffic Flow Theory. Physical Review E **53**, 4655-4672

Najneen, F., Hoque, K.S., Mahmood, S.M.S., Rahman, S. & Sharmin, M. Traffic congestion due to unplanned activities. Bangladesh research publications Journal, 4(2), pp. 185–197, 2010

National Renewable Energy Laboratory Project Draft Final Report for the Period August 1, 2012, through March 31, 2014, "Data Collection, Testing and Analysis of Hybrid Electric Trucks and Buses Operating in California Fleets." ARB Agreement Number 11-600, NREL Contract Number FIA-12-1763, April 15, 2014.

Nelson, P., D.D. Bui. Moreover, A. Sopasakis (1997). A Novel Traffic Stream Model deriving from a Bimodal Kinetic Equilibrium. Proceedings of the IFAC conference, 799-804.

Rastorfer, Robert L., \Toll Plaza Concepts", presentation at ASCE Fall Conference, Houston, 2004.

Zaynab Yagi, 2015.Traffic congestion costs one million dollar, Assafir journal.

Ruben Nuredini, jasmine Ramadani (2011). Performance measurement model of pay-toll system.

Sundarapandian, V. (2009). 7. Queueing Theory: Probability, Statistics and Queueing Theory. PHI Learning. ISBN 8120338448

Walsh, M. (1990) Global Trends in Motor Vehicle use and Emissions. In Annual Review of Energy, Volume 15, pp. 217-243

Wand, Liu and Montgomery, "A simulation model for motorway merging behavior", institute for transport studies, university of Leeds.

Scoring Vulnerabilities After Seeing a Chained Vulnerability Demonstration

Nikki Robinson
Capitol Technology University
nerobinson@captechu.edu

Abstract— The general problem was the NIST SP 800-40r3 (Souppaya & Scarfone, 2013) or the CVSS (FIRST, 2018a) did not provide enough information to prioritize vulnerability remediation. The specific problem was CVSS severity rankings were specific to individual vulnerabilities, which limited organizations to remediate vulnerabilities based on the potential downstream impact to other systems (Franklin, Wergin, & Booth, 2014). The purpose of this quantitative study was to use a pre-test / pro-test experiment to compare how cybersecurity professionals in the USMC rate vulnerabilities before and after seeing examples of vulnerability chaining using the CVSS calculator. The research question was, what score would cybersecurity professionals in the USMC give individual vulnerabilities before and after seeing vulnerabilities used in combination to create a more severe cyberattack? The research method used a quasi-experimental method with a pre-test / post-test design to identify how vulnerabilities would be scored before and after seeing a chained vulnerability demonstration. The results of the vulnerability scores were compared between the control and treatment groups, as well as the CVSS scores provided in versions 2.0 and 3.0 for each vulnerability. Participants from the control group changed two vulnerabilities from a Medium score to a High score; CSRF (from 7.5 to 9.0) and XSS (8.3 to 9.0). The treatment group did not change any vulnerability scores in a statistically significant manner, but the researcher found this was due to the overall higher scores for each vulnerability.

Keywords—vulnerability, chaining, NIST, scores

I. INTRODUCTION

This study explored the importance of Medium vulnerabilities, and how vulnerabilities were used in combination to create an attack as detrimental as a high or critical. Since the detrimental cyberattacks against large organizations, including the Equifax breach of 2017 (Berghel, 2017) and the Office of Personnel Management (OPM) breach of 2015 (Harvey & Evans, 2016), government agencies and businesses must be hyper vigilant about remediating vulnerabilities to ensure the protection of sensitive data. With the increased threat of cybersecurity attacks (Hammond, 2016), organizations may only have time to focus on the remediation of Critical and High vulnerabilities. If an organization opts out of addressing vulnerabilities which were classified at a lower level, were they more susceptible to a cyberattack? Should organizations create and maintain patch management solutions for Low and Medium exploitable vulnerabilities?

The National Institute of Standards and Technology (NIST) provided guidance for government agencies to create patch management and configuration management documentation, but it does not sufficiently deliver guidance for remediating vulnerabilities. However, the NIST SP 800-40r3 was not the only guide used to create patching strategies. The Common Vulnerability Scoring System (CVSS) is used to rate the severity of vulnerabilities but has moved from version 2.0 to version 3.0 (FIRST, 2018a). When versions CVSS 2.0 and CVSS 3.0 were compared, vulnerability scores for some vulnerabilities changed from *Medium* to *High*, and vice versa (FIRST, 2018b). This creates difficulty in patch and risk mitigation strategies as old vulnerabilities can be re-classified at a higher level, which lead to the reason for this study.

To address the potential gaps in knowledge, this study discovered if individuals rank *Low* and *Medium* vulnerabilities differently after knowledge of a chained attack. The introduction explored the explanation of why this research was important, general and specific problem statements, purpose, significance, and nature of the study. Further into this paper the research questions, the theoretical framework, definitions, assumptions, and scope of the study will be discussed.

II BACKGROUND

The general problem is the NIST SP 800-40r3 (Souppaya & Scarfone, 2013) or the CVSS (FIRST, 2018a) does not provide enough information to prioritize vulnerability remediation. Organizations were not given enough detail for each vulnerability to create a proper patch management plan to secure their environment. Both documents left prioritization and mitigation of vulnerabilities to the organization, which could leave critical applications or legacy systems vulnerable to cyberattacks.

The specific problem is the CVSS severity rankings are specific to individual vulnerabilities, which limited organizations to remediate vulnerabilities based on the potential downstream impact to other systems (Franklin, Wergin, & Booth, 2014). *Low* or *Medium* vulnerabilities can be used together to create a more sophisticated and harmful attack (FIRST, 2018a). Organizations may not be aware of the importance of patching or remediating *Low* and *Medium* vulnerabilities. By gaining more knowledge about vulnerability chaining, and how this was used to compromise systems, organizations will be better able to prioritize vulnerability remediation and create a more secure environment.

The purpose of this quantitative study is to use a pre-test / pro-test quasi-experiment to compare how cybersecurity professionals in the USMC rate vulnerabilities before and after seeing examples of vulnerability chaining using the CVSS calculator. This study also showed how vulnerability chaining was used with what the CVSS scoring system classify as *Low* and *Medium* vulnerabilities to provide more complex cyberattack. The intention of the pre-test measurement was to discover if individuals understood the basic CVSS scores and accurately scored *Low* and *Medium* vulnerabilities. The intention of the post-test measurement was to find out if, after seeing a chained attack, the individuals scored those vulnerabilities differently. In this quasi-experiment, the researcher hoped to find out if examples of vulnerability chaining would change the overall score of an individual vulnerability.

A Research Questions

The question this research sought to answer was, what was the score that cybersecurity professionals in the USMC gave individual vulnerabilities before and after seeing vulnerabilities used in combination to create a more severe cyberattack? The hypothesis was the CVSS score that USMC cybersecurity professionals assigned to vulnerabilities after seeing chained attacks will be statistically different than a control group not exposed to the chained attacks. The null hypothesis was the CVSS score that USMC cybersecurity professionals assigned to vulnerabilities after seeing chained attacks will not be statistically different than a control group not exposed to the chained attacks. Through identifying how these individuals rank vulnerabilities, there were several qualitative research questions, which could be asked, based on the results.

Research Question: Will cybersecurity professionals in field operations of the Marine Corps rank vulnerabilities higher, the same, or lower after seeing how vulnerabilities can be chained together?

Hypothesis: The CVSS score that Marine Corps cybersecurity professionals assign to vulnerabilities after seeing chained attacks will be statistically different than a control group not exposed to the chained attacks.

(Null) Hypothesis: The CVSS score that Marine Corps cybersecurity professionals assign to vulnerabilities after seeing chained attacks will not be statistically different than a control group not exposed to the chained attacks.

I. Data Analysis and procedures

The set of procedures for this quasi-experimental quantitative study were outlined to include the selection of participants and the control group. Before the control group or treatment group were surveyed, a pilot study took place to ensure validity and reliability of the quasi-experimental research design. The participants were selected at random for the control group and the experiment using a link provided to the USMC POC to distribute. Each participant was assigned a random number by a research assistant to ensure the anonymity of the individuals. Half of the sample was chosen at random to participate in the control group based on the AB text function in SurveyMonkey. The remaining half of participants were used in the experiment.

From both the control group and the experimental group, several pieces of data were analyzed, as follows. From the control group, it was important to find out how cybersecurity professionals ranked vulnerabilities with (or without) knowledge of chained vulnerability scenarios. The demographic information gathered from the participants provided insight into if knowledge of vulnerability chaining determines the score of vulnerabilities. Another piece of data to analyze was how much experience everyone had in cybersecurity and vulnerability management. Examining the control group was important before looking into the group which receives the treatment.

Once the survey results were analyzed from the control group, the treatment group results were examined. The most important component of the study was to see if scores changed between the pre-test and post-test scores from participants. This was done using a chart to show how individuals scored vulnerabilities before and after the demonstration. After determining if vulnerabilities were scored higher, lower, or the same, the length of time to complete the survey was inspected. It was possible the length of time to complete the survey meant the participants were unsure of how to answer or did not understand the question.

Inferential statistics were used to test the hypothesis, and the sample was identified using the statistics related to demographics. An inferential statistic review consisted of looking at the independent variable related to the dependent variable and using a One Sample t Test (Kent State University, 2018). This type of test looked at the statistical difference between zero and a change score (Kent State University, 2018). This test helped to determine if there was a change in score compared to the original measurements (scores for vulnerabilities provided by CVSS) (Kent State University, 2018).

I. chosen Vulnerabilities for experiment

The third and fifth pages of the survey were the pre-test and post-test questions and included the same in format and content. There was a description above each of these pages, which included how CVSS scores vulnerabilities ranged from *None* (0.0), *Low* (0.1 – 3.9), *Medium* (4.0 – 6.9), *High* (7.0 – 8.9), or *Critical* (9.0 – 10.0). The description also requested the participants to score vulnerabilities in a decimal format and provided an example for how to correctly input responses. Nine vulnerabilities were presented including a description directly from the OWASP website. The intention of providing the description was to ensure each participant understood fully which vulnerability to score. After each description, the participant was provided with a prompt to score the vulnerability based on the given description, as well as their own experience with vulnerability management.

Each of the nine vulnerabilities was chosen based on the inconsistencies of vulnerability score between versions 2.0 and 3.0 of CVSS, as well as any iterations of the vulnerability where it was scored as a *Medium*. The nine vulnerabilities chosen included Server-Side Request Forgery (SSRF), Cross-Site Request Forgery (CSRF), Carriage Return Line Feed (CRLF) Injection, Deserialization, HTTPOnly flag, Cross-site Scripting (XSS), Remote File Inclusion (RFI), SQL Injection (SQLi), and Authentication Bypass. Depending on the software or operating system affected by these vulnerabilities, the CVSS score was dissimilar. Each of these vulnerabilities was described, CVE number and description provided, as well as a possible CVSS score.

A. *SSRF*

The first vulnerability presented was SSRF, which was the potential for an attacker to make custom requests to a web server on an internal network (Särud, 2018). One example of a SSRF vulnerability was CVE-2018-1999039 which was released in August of 2018 (NIST, 2018i). This vulnerability was classified with a base score of 4.3 (*Medium*) in CVSS 3.0 and a base score of 4.0 (*Medium*) in CVSS 2.0 (NIST, 2018i). This specific CVE is for an SSRF which existed in plugin version 2.0.1 for Jenkins Confluence Publisher (NIST, 2018i). Depending on the software affected, CVE's related to SSRF on the NIST website were classified *Medium*, *High*, or *Critical*.

B. *CSRF*

The second vulnerability presented was CSRF, which allowed an attacker to log the victim in to a system with the attacker's credentials (Särud, 2018). One example of a CSRF vulnerability was CVE-2018-13401 which was released in December of 2018 (NIST, 2018f). This vulnerability was classified with a base score of 6.1 (*Medium*) in CVSS 3.0 and a base score of 5.8 (*Medium*) in CVSS 2.0 (NIST, 2018f). This specific CVE was for a CSRF vulnerability which existed in Jira versions before 7.13.1 and allowed attackers to obtain the CSRF token through a redirect vulnerability (NIST, 2018f). Depending on the software affected, CVE's related to CSRF on the NIST website were classified *Medium*, *High*, or *Critical*.

C. *CRLF*

The third vulnerability presented was CRLF Injection, which allowed an attacker to inject a header to grant internal servers the permissions to deploy other systems via a callback (CWE, 2019a). One example of a CRLF Injection was CVE-2017-7528 which was released in August of 2018 (NIST, 2018c). This vulnerability was classified with a base score of 6.5 (*Medium*) in CVSS 3.0 and a base score of 3.3 (*Low*) in CVSS 2.0 (NIST, 2018c). This specific CVE was for Ansible Towers which had Red Hat Engine 5 and was vulnerable to CRLF attacks (NIST, 2018c). Depending on the software affected, CVE's related to CRLF on the NIST website were classified *Low*, *Medium*, or *High.*

D. *Deserialization*

The fourth vulnerability presented was Deserialization, which was when an application deserialized untrusted data without verifying the information was valid (CWE, 2019b). One example of a Deserialization vulnerability was CVE-2016-9585, which was released in March of 2018 (NIST, 2018b). This vulnerability was classified with a base score of 5.3 (*Medium*) in CVSS 3.0 and a base score of 2.6 (*Low*) in CVSS 2.0 (NIST, 2018b). This CVE was for Red Hat JBoss EAP version 5, which was vulnerable to deserialization in the JMX endpoint (NIST, 2018b). This type of vulnerability could result in a denial of service attack against the machine (NIST, 2018b). Depending on the software affected,

CVE's related to Deserialization on the NIST website were classified *Medium*, *High*, or *Critical*.

E. HTTPOnly Flag

The fifth vulnerability presented was the HTTPOnly flag, which allowed attackers to obtain sensitive information through access to cookies (NIST, 2018a). One example of an HTTPOnly flag vulnerability was CVE-2014-9635, which was released in September 2017 (NIST, 2018a). This vulnerability was classified with a base score of 5.3 (*Medium*) in CVSS 3.0 and a base score of 5.0 (*Medium*) in CVSS 2.0 (NIST, 2018a). This CVE was for any version of Jenkins before 1.586 which did not set a Set-Cookie header when run on Tomcat 7.0.41 (NIST, 2018a). Depending on the software affected, CVE's related to HTTPOnly flags on the NIST website were classified *Medium*, *High*, or *Critical*.

F. XSS

The sixth vulnerability was XSS, which was able to steal cookies using this vulnerability. An example of an XSS vulnerability is CVE-2018-17952, which was released in December 2018 (NIST, 2018h). This vulnerability was classified with a base score of 6.1 (*Medium*) in CVSS 3.0 and a base score of 4.3 (*Medium*) in CVSS 2.0 (NIST, 2018h). This CVE was for an XSS vulnerability in eDirectory software prior to version 9.1 Service Pack (SP) 2 (NIST, 2018h). Depending on the software affected, CVE's related to XSS on the NIST website were classified *Not Available*, *Medium*, or *High*. The most surprising find about this vulnerability were the amount of XSS CVE's which did not have a CVSS version 2.0 or 3.0 score, considering XSS made OWASP's Top 10 list in 2017 (OWASP, 2018a).

G. RFI

The seventh vulnerability presented was RFI, which allowed a directory to be loaded as a file into a share (Kure, 2015). An example of an RFI vulnerability was CVE-2018-11101, which was released in May of 2018 (NIST, 2018e). This vulnerability was classified with a base score of 6.1 (*Medium*) in CVSS 3.0 and a base score of 4.3 (*Medium*) in CVSS 2.0 (NIST, 2018e). This CVE mentioned the use of two vulnerabilities, including XSS and using RFI to inject HTML code as a message (NIST, 2018e). Depending on the software affected, CVE's related to RFI on the NIST website were classified *Not Available*, *Medium*, or *High*.

H. SQL Injection

The eighth vulnerability presented was SQLi, which allowed an attacker to execute SQL commands on a database to read data from tables (NIST, 2018d). An example of a SQLi vulnerability is CVE-2018-11065, which was released in August of 2018 (NIST, 2018d). This vulnerability was classified with a base score of 4.3 (*Medium*) in CVSS 3.0 and a base score of 4.0 (*Medium*) in CVSS 2.0 (NIST, 2018d). This CVE was related to a component of RSA Archer, called WorkPoint, and was only vulnerable on versions prior to 6.4.0.1 (NIST, 2018d). Depending on the software affected, CVE's related to SQLi on the NIST website were classified *Medium*, *High*, or *Critical*.

I. Authentication Bypass

The ninth vulnerability presented was an Authentication Bypass, which was accomplished by using a modified URL parameter, manipulating a form, or counterfeiting sessions from the user (OWASP, 2018b). An example of an Authentication Bypass vulnerability was CVE-2018-1650, which was released in December of 2018 (NIST, 2018g). This vulnerability was classified with a base score of 5.5 (*Medium*) in CVSS 3.0 and a base score of 2.1 (*Low*) in CVSS 2.0 (NIST, 2018g). This CVE was specific to IBM QRadar SIEM versions 7.2 and 7.3, which allowed attackers to bypass authentication the administrator had configured (NIST, 2018g). Depending on the software affected, CVE's related to SQLi on the NIST website were classified *Medium*, *High*, or *Critical*. Table 2 shows how each vulnerability was scored in CVSS version 2.0 versus CVSS version 3.0.

TABLE I.

Vulnerability	**CVSS v2.0 Versus v3.0 Scores**	
	CVSS v2.0	***CVSS v3.0***
SSRF	4.3	4.0
CSRF	6.1	5.8
CRLF	6.5	3.3
Deserialization	5.3	2.6
HTTPOnly	5.3	5.0
XSS	6.1	4.3
RFI	6.1	4.3
SQLi	4.3	4.0
Authentication Bypass	5.5	2.1

After the participants were shown the pre-test survey page, which requested vulnerability scores for each of the nine vulnerabilities, the treatment group received a page containing possible chained vulnerability attacks. These attacks were chosen based on the potential for them to exist in real-world scenarios. Each chained vulnerability example contained a reference to someone who either used these chained vulnerability attacks or proposed the potential to exploit. Figure 1 displays how each of the chained vulnerability examples worked, which included the vulnerabilities presented in the pre-test and post-test questions.

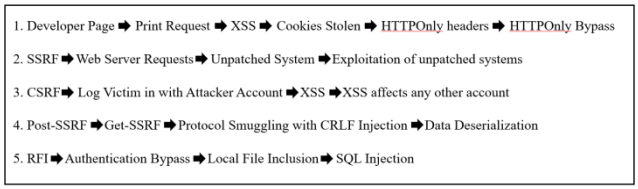

Fig 1. Vulnerability chain examples used in the experiment.

II. RESULTS

A. Pilot Results

The initial pilot participant noted some issues with understanding how and why the vulnerabilities were chosen. The individual believed the vulnerabilities were more like vulnerability categories, instead of specifically chosen vulnerabilities with a CVE ID. The individual also noted the researcher used OWASP definitions for vulnerabilities, where the individual thought STIG references would be more appropriate. The researcher politely disagreed with the individual and explained this in a follow-up e-mail.

The vulnerabilities were chosen because each one specifically fit into a vulnerability chaining example, not because they were based on any one technology or vulnerability in a system. The reason OWASP definitions were used were based on its wide usage within the cybersecurity community. OWASP was also used because several of the vulnerabilities were included in the OWASP Top Ten list (as mentioned in Chapter 3). Once the participant received this information, they agreed the definitions and vulnerabilities were correct and was satisfied with the three changes the researcher made to the survey.

The other participants did not provide additional feedback after taking the survey. As the researcher did receive helpful and descriptive information from one participant, this helped to shape the study. The pilot participants were asked to complete the survey within five business days and three of the participants complied. On the closing date the researcher sent a final note to Dr. Letteer with detailed instructions on the changes to the survey. The link to the survey was also provided to Dr. Letteer for distribution to the USMC cybersecurity groups.

B. *Control Group*

A total of 3 participants received the survey without the treatment, but only 2 participants scored both sets of vulnerabilities. The control group took an average of 20 minutes to complete the survey without the treatment. Table 3 contains the results of the control group pre-test and post-test mean scores. The vulnerabilities which saw a statistical difference of +/- 1 were CSRF and HTTPOnly. Authentication Bypass was the sole vulnerability which received a lower score, though it would not be considered statistically significant per the CVSS scale. The standard deviation was included to provide context between the pre-test and post-test mean scores.

Table ii.

Vulnerability	Control Group Results		
	Control Pre-Test Mean	**Control Post-Test Mean**	**Std. Deviation**
SSRF	8.7	9.2	9
CSRF	7.5	9	8.28
CRLF	8.2	8.2	0
Deserialization	9.5	9.6	9.57
HTTPOnly	7.6	8.7	8.19
XSS	8.5	9	8.75
RFI	9.6	9.7	9.7
SQLi	9	9	0
Authentication Bypass	9.2	9	9.12

C. *Treatment Group*

A total of 7 participants received the survey with the treatment, but only 6 completed scoring of both sets of vulnerabilities. The treatment group took an average of 27 minutes to complete the survey with the treatment. Table 4 contains the results of the treatment group pre-test and post-test mean scores. The vulnerabilities which saw a statistical difference of +/- 1 were CSRF and XSS. HTTPOnly and RFI received a lower score, though it would not be considered statistically significant per the CVSS scale. The standard deviation was included to provide context between the pre-test and post-test mean scores.

Table iii.

Vulnerability	**Treatment Group Results**		
	Treatment Pre-Test Mean	*Treatment Post-Test Mean*	*Std. Deviation*
SSRF	8.5	8.5	0
CSRF	7.4	8.6	8.06
CRLF	6.3	6.5	6.44
Deserialization	7.8	7.9	7.89
HTTPOnly	3.6	3.1	1.83
XSS	7.8	8.9	8.84
RFI	8.3	7.9	8.29
SQLi	9.1	9.2	9.16
Authentication Bypass	8.1	8.2	8.15

D. *Mean – Control Group*

Table 4 displays the variance between CVSS version 2.0, version 3.0, and control group pre-test and post-test scores. These scores show the wide variance not only between CVSS version 2.0 and 3.0, but also how the participants scored the vulnerabilities with only their knowledge and experience. While the pre-test and post-test scores are statistically similar, vulnerabilities show a statistically significant score for all vulnerabilities between CVSS versions and the control groups responses.

Table iv.

Vulnerability	**Control Group Results - Mean**			
	CVSS v2.0	*CVSS v3.0*	*Control Pre-Test Mean*	*Control Post-Test Mean*
SSRF	4.3	4.0	8.7	9.2
CSRF	6.1	5.8	7.5	9
CRLF	6.5	3.3	8.2	8.2
Deserialization	5.3	2.6	9.5	9.6
HTTPOnly	5.3	5.0	7.6	8.7
XSS	6.1	4.3	8.5	9
RFI	6.1	4.3	9.6	9.7
SQLi	4.3	4.0	9	9
Authentication Bypass	5.5	2.1	9.2	9

E. *Mean – Treatment Group*

Table 5 displays the variance between CVSS version 2.0, version 3.0, and treatment group pre-test and post-test scores. These scores show a wide variance not only between CVSS version 2.0 and 3.0, but also how the participants scored the vulnerabilities with only their knowledge and experience. While the pre-test and post-test scores are statistically similar on all but CSRF and XSS, vulnerabilities show a statistically different score for all vulnerabilities between CVSS versions and the control groups responses.

Table v.

Vulnerability	**Treatment Group Results - Mean**			
	CVSS v2.0	*CVSS v3.0*	*Trtmt Pre-Test Mean*	*Trtmt Post-Test Mean*
SSRF	4.3	4.0	8.5	8.5

Vulnerability	Treatment Group Results - Mean			
	CVSS v2.0	*CVSS v3.0*	*Trtmt Pre-Test Mean*	*Trtmt Post-Test Mean*
CSRF	6.1	5.8	7.4	8.6
CRLF	6.5	3.3	6.3	6.5
Deserializatio n	5.3	2.6	7.8	7.9
HTTPOnly	5.3	5.0	3.6	3.1
XSS	6.1	4.3	7.8	8.9
RFI	6.1	4.3	8.3	7.9
SQLi	4.3	4.0	9.1	9.2
Authentication Bypass	5.5	2.1	8.1	8.2

III. Conclusions

To address the research question and hypothesis, the interpretation of the findings will be discussed. Findings will include any relation between demographic information and vulnerability scores, relation to participants vulnerability scores and CVSS scores, as well as any difference between the control and treatment groups. Of major interest to the paper, will be the analysis of control and treatment groups, to find if vulnerability scores changed due to the chained vulnerability examples. The researcher will also discuss if the evidence collected was bound or unbound, and if the data confirms or contradicts the research question.

The research question asked if USMC personnel would score vulnerabilities differently or the same after reviewing a demonstration of chained vulnerabilities. While reviewing the differences in pre-test and post-test scores, the control group scored two of the nine vulnerabilities the same in both sets of questions, while the other vulnerability scores changed. The vulnerabilities scores which did not change were CRLF (8.2) and SQL injection (9). The average vulnerability score for the control group changed two vulnerabilities from a *Medium* score to a *High* score according to the CVSS calculator; CSRF (from 7.5 to 9.0) and XSS (8.3 to 9.0). However, one vulnerability score went down on the second round of scoring, the Authentication Bypass was changed from a 9.2 to a 9.0. Results from the control group were inconsistent, but the only statistically significant changes were shown in the CSRF and XSS vulnerabilities. This could just mean that the participants meant to

The hypothesis noted that USMC cybersecurity professionals would score the vulnerabilities in a statistically different way than the control group, which did not see the chained vulnerability demonstration. There were four vulnerabilities which saw statistically significant changes in scores between the control group and the treatment group. CRLF, Deserialization, HTTPOnly, and RFI vulnerabilities were scored at a +/-1 between the control and the treatment groups. This showed a difference in how vulnerabilities were scored after seeing a chained vulnerability attack, but the most fascinating part was the vulnerabilities were scored lower in the treatment group. The treatment group scored vulnerabilities overall lower than the control group. So, while the vulnerability scores themselves from the treatment group were not statistically different, compared to the control group it was a significant difference.

The most interesting finding of this study was the increased scores that participants gave to all vulnerabilities, except for the HTTPOnly flag. This was interesting since CVSS scored this vulnerability +/- 2 points higher than the average participant. It is possible this vulnerability is not as well known, and therefore could be not well understood in the community. The vulnerability chaining example did not increase the score of this vulnerability but decreased in the post-test treatment group. But the reasoning for this remains a mystery, as there is no indication in the data why the individuals chose to score this vulnerability lower after the demonstration. One could deduce the participants found this vulnerability less severe after seeing how it was used in a chained attack.

IV Future work

While the analysis determined that the treatment group only changed one vulnerability in a statistically significant manner, this leads to many other questions the researcher would like to ask. The first question is why these individuals did not score the other vulnerabilities higher, or whether the treatment did not affect their initial higher scores of the vulnerabilities. A qualitative study could be done with the same sample, to find if there is a correlation between the training they receive and the ability to score vulnerabilities higher. Another quantitative study could be done to show the exact CVE numbers and find if the individuals would be able to accurately score the vulnerabilities based on explicit technical detail.

Another possibility for future research is to perform this experiment with either a government or private industry organization. It could be very interesting to find out how other participants would score vulnerabilities, and if their training and education background would change their answers. The USMC clearly has some excellent documentation and training material, and it seems like their staff is well trained on vulnerability scoring. It would also be helpful to find a larger sample to conduct this experiment on. If the questions were shortened, it is possible more people would be willing to complete the survey.

It could also be interesting to choose another set of vulnerabilities and show new vulnerability chaining examples. The intention of this research was to take relatively well-known vulnerabilities to score, but it may be more interesting to choose lesser known vulnerabilities to see how the scores would change. This could lead to further research as participants may need more information on the vulnerabilities and could potentially be done using a qualitative method. It could provide more concrete evidence as to why individuals score vulnerabilities, and not strictly the scoring numbers.

References

Easttom, C. (2018, March). The Role of Weaponized Malware in Cyber Conflict and Espionage. In *ICCWS 2018 13th International Conference on Cyber Warfare and Security* (p. 191). Academic Conferences and publishing limited.

Berghel, H. (2017). Equifax and the latest round of identity theft roulette. *Computer; New York*, *50*(12), 72-76. doi: http://dx.doi.org.login.captechu.edu:2048/10.1109/MC.2017.4451227

Common Weakness Enumeration. (2019a). *CWE-93: Improper neutralization of CRLF sequences ('CRLF injection')*. Retrieved from http://cwe.mitre.org/data/definitions/93.html

Common Weakness Enumeration (2019b). *CWE-502: Deserialization of untrusted data*. Retrieved from http://cwe.mitre.org/data/definitions/502.html

FIRST (2018a). *Common vulnerability scoring system v3.0: Examples*. Retrieved from https://www.first.org/cvss/specification-document

FIRST (2018b). *Common vulnerability scoring system v3.0: Specification document*. Retrieved from https://www.first.org/cvss/examples

Franklin, J., Wergin, C., & Booth, H. (2014). National Institute of Standards and Technology (2014, April). *CVSS Implementation Guidance*. NIST Interagency Report 7946. doi: http://dx.doi.org/10.6028/NIST.IR.7496

Hammond, B. (2016). DHS official: Cyber threat data should be public good more than profit maker. *Cybersecurity Policy Report*. Retrieved from https://search-proquest-com.login.captechu.edu:2443/docview/1773928842?accountid=44888

Harvey, S., & Evans, D. (2016). *Defending against cyber espionage: The US office of personnel management hack as a case study in information assurance*. Paper presented at the 2016 National Conference on Undergraduate Research, University of North Carolina Asheville, Asheville, NC.

Kent State University (2019, February 1). *SPSS tutorials: One sample t test*. Retrieved from https://libguides.library.kent.edu/SPSS/OneSampletTest

National Institute of Standards and Technology (2018a). CVE-2014-9635 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2014-9635

National Institute of Standards and Technology (2018b). CVE-2016-9585 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2016-9585

National Institute of Standards and Technology (2018c). CVE-2017-7528 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE2017-7528

National Institute of Standards and Technology (2018d). CVE-2018-11065 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2018-11065

https://nvd.nist.gov/vuln/detail/CVE-2018-11101
National Institute of Standards and Technology (2018f). CVE-2018-13401 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2018-13401

National Institute of Standards and Technology (2018g). CVE-2018-1650 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2018-1650

National Institute of Standards and Technology (2018h). CVE-2018-17952 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2018-17952

National Institute of Standards and Technology (2018i). CVE-2018-1999039 detail. Retrieved from https://nvd.nist.gov/vuln/detail/CVE-2018-1999039

Open Web Application Security Project (2018a). OWASP top 10 – 2017.Retrieved from https://www.owasp.org/images/7/72/OWASP_Top_10-2017_%28en%29.pdf.pdf

Open Web Application Security Project (2018b). Testing for bypassing authentication schema (otg-authn-004). Retrieved from https://www.owasp.org/index.php/Testing_for_Bypassing_Authentication_Schema_(OTG-AUTHN-004)

Särud, L. (2018, February 6). *Do not dismiss the small vulnerabilities!* [Web blog post]. Retrieved from https://blog.detectify.com/2018/02/06/small-vulnerabilities/

Souppaya, M., & Scarfone, K. (2013, July). National Institute of Standards and Technology. *Guide to enterprise patch management technologies*. NIST Special Publication 800-40; Revision 3. Retrieved from http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-40r3.pdf

Gene Selection and Classification Using Quantum Moth Flame Optimization Algorithm

1. Ali Dabba[a,c,d]
2. Abdelkamel Tari[a,e]
3. Samy Meftali[b,d]

[a]*Faculty of Sciences,Computer Science Department, Abderrahmane Mira University,Bejaia,Algeria*
[b]*University of Lille, France*
[c]*Faculty of Mathematics and Computer Science, Computer Science Department, Mohamed Boudiaf University, M'sila, Algeria.*
[d]*Research center in Computer Science, Signal and Automatic Control of Lille – CRIStAL*
[e]*Laboratory of Medical Computing - LIMD*

***ABSTRACT* - In this paper, we present a new swarm intelligence algorithm for gene selection called quantum moth flame optimization algorithm (QMFOA), which based on hybridization between quantum computation and moth flame optimization algorithm (MFOA). The purpose of QMFOA is to identify a small gene subset that can be used to classify samples with high accuracy. The QM-FOA has a simple two-phase approach, the first phase is a pre-processing that uses to address the difficulty of high-dimensional data, which measure the redundancy and the relevance of the gene, in order to obtain the relevant gene set. The second phase is hybridization among MFOA, quantum computing, and support vector machine (SVM) with leave-one-out cross-validation (LOOCV), in order to solve the gene selection problem. The main objective of the second phase is to determine the best relevant gene subset of all genes obtained in the first phase.**

In order to assess the performance of the proposed QMFOA, we test it on six Microarray datasets. Experimental results show that QMFOA provides great classification accuracy in comparison to some known algorithms.

KEYWORDS* - *Genes expression, Feature Selection, Moth Flame Optimization, Algorithm Quantum Computing, Microarray Data, Cancer Classification, Bio-inspired, Algorithms Molecular, Biology Optimization, Algorithms, Evolutionary, Algorithms Swarm, Intelligence.

1. INTRODUCTION

Gene selection is a branch of feature selection, which establishes an evident approach to reducing dimensionality and over-fitting [17]. The main task of gene selection is to find the best subset of genes from all possible choices by filtering out irrelevant, redundant and noisy genes [25]. To achieve good classification accuracy, it is important to choose the most pertinent genes that are necessary and sufficient to describe the target concept, like gives some aspects of functional genomics. In addition, to find an optimal small set of relevant genes has been proven to be an NP-complete problem [5, 7].

In literature, several gene selection methods have been proposed and can be organized into three categories including filter, wrapper, and embedded methods [18, 9]. Filter methods utilize essentially the general statistical properties of the training data at hand without using any learning algorithm. Although these methods are fast but have rather poor performance. In contrast, the wrapper methods select a set of discriminatory features by using a predetermined learning algorithm. The interest of these methods is that the chosen subset is perfectly adapted to the classifier. How- ever, the wrapper methods are more costly in computational time because each evaluation of a feature subset requires a training model, in which the computational complexity depends on the complexity of the learning model used [8]. Embedded methods are similar to wrapper approaches by combining the exploration process with a learning algorithm [10], which are an extension of wrapper approaches and undertake feature selection in the process of classifier training. The advantage of these methods is that the classifier provides important information that guides the search, which makes these methods more efficient than wrapper methods.

In recent years, quantum computing has been proposed in the literature [19, 20] as a more effective technique than classical computing. In other words, quantum physics has been used to build a new kind of computers, called quantum computers [20]. Unlike classical computers that deal with binary digits (bits), the basic unit a quantum bit (Q-bit), in addition the usual f0f and f1f states, a Q-bit can also in any superposition of these two states [13]. Therefore, the best suggestion right now is to use quantum algorithms and apply them to classical computers.

Over the past few years, many algorithms have been proposed to solve gene selection using quantum fields (quantum computing). Among them, Cluster QGA have been proposed by [23], which uses clustering to choose a small set of non-redundant representative genes and then applies the Quantum Genetic Algorithm to define a minimal set of non-redundant and relevant genes. The authors in [26] have proposed an approach called binary quantum-behaved particle swarm optimization (BQPSO). This approach coupling between PSO, quantum computing, and support vector machine (SVM) with leave-one-out cross-validation to solve gene selection, which is a discretized version of the original QPSO for binary 0-1 optimization problems. In addition, the $GQA_{SY\ M}$ [1] has been proposed as a hybrid approach between the Genetic Quantum Algorithm and the Support Vector Machines classifier to gene selection and classification of Microarray Data. The main goal of this algorithm is to identify a small subset of genes that could be used to separate two classes of samples with high accuracy.

In this work, we propose a new algorithm called Quantum Moth Flame Optimization Algorithm (QMFOA), in order to find the best gene subset to provide high classification accuracy to cancer Microarray data. The QMFOA inherits parallelism, decentralization, and cooperation of swarm intelligence algorithm (bio-inspired algorithm, specifically MFOA) to solve the gene selection problem. For solve this problem, the QMFOA uses a hybrid model that uses several techniques: Quantum field (Quantum computing), Moth Flame Optimization Algorithm (MFOA), Mini- mum Redundancy-Maximum Relevance (mRMR), and a Support Vector Machine (SVM) with Leave One Out Cross Validation (LOOCV).

In order to prove the advantages of proposed QMFOA, we have tested QMFOA on six well-known datasets issued of Microarray experiments treating cancer and compared our results with several recently published algorithms in the literature. The experimental results have shown that QMFOA can achieve better performance of classification accuracy with a competitive number of genes selected i.e., it is able to provide a minimum number of genes to obtain the highest classification accuracy for solving the gene selection problem in both binary and multi classes.

The remainder of the paper is organized as follows: Section 2 presents our proposed QMFOA approach to gene selection. The experimental results and discussions are included in Section 3. Finally, the conclusion is given in Section 4.

2. THE PROPOSED ALGORITHM FOR THE GENE SELECTION PROBLEM

In this section, we propose a new algorithm called Quantum Moth Flame Optimization Algorithm (QMFOA) for gene selection and classification of high dimensional Microarray data. This work is based on a hyridizing Moth- Flame Optimization Algorithm (MFOA) with concepts resulting from the quantum field to provide solutions for a gene selection problem. However, the QMFOA purpose is to select small samples of informative genes amongst thousands of them.

The principle of the proposed algorithm consists of a two-phase approach. In the first phase, instead of to use the full set of available genes, we use preprocessing to select a relatively smaller set of non-redundant and relevant genes and that passed on to the second phase of QMFOA for the effective selection of a minimal set of informative genes. In order to guarantee this, we start by normalization of Microarray data with the Min-Max method that can guarantee a stable convergence of weights and biases [14]. Furthermore, we use the statistical technique of Minimum Redundancy-Maximum Relevance (mRMR) to measure the relevance and redundancy of selected genes, in order to reduce the high number of genes by eliminating genes redundancy [21].

The second phase of QMFOA is applied to the set of *d* representative genes that were obtained in the first phase of QMFOA in order to find a minimal subset of the relevant genes (.i.e, the maximum number of genes in a moth (individual)). Like any bio-inspired algorithms, this algorithm is based on a population of solutions that is preserved through several generations, which seeks the best-fitted solution to the gene selection problem, evaluating the gene subset of those included in the current population. The fundamental idea of this phase is to combine the MFOA and quantum fields with the SVM classifier. The fitness function uses the SVM classifier with the Leave One Out Cross Validation (LOOCV) method and the percentage of genes that are not selected, which is applied in order to evaluate and validate the provided solutions. The main goal of QMFOA is to select a high accuracy genes subset that includes a smaller number of genes. Finally, the better gene subset obtained by QMFOA will be evaluated using the SVM classifier.

2.1 Representation of Candidate Solutions

For QMFOA, the moth (individual) represents a gene subset of the *dsubset*: set of genes; obtain maximum number *d* of genes in an individual.
The moth population containing *n* and the number *d* of Q-bits. it is represented as $QM = \{Qm_1, Qm_2, \ldots, Qm_d\}$,

Where, Qm_i (i = 1, 2,..., n) is the i^{th} moth. Each Qm_i quantum moth represents as follows (Eq.1):

$$Qm_1 = \begin{vmatrix} \cos(\theta_{i,1}) & \cos(\theta_{i,2}) & \cos(\theta_{i,3}) & \ldots\ldots & \cos(\theta_{i,d}) \\ \sin(\theta_{i,1}) & \sin(\theta_{i,2}) & \sin(\theta_{i,3}) & \ldots\ldots & \sin(\theta_{i,d}) \end{vmatrix}$$

Where, d is the number of Q-bits used in each quantum moth's representation, $\theta_{i,k}$ (k = 1, ..., d) with $\theta_{i,k} \epsilon [0, \frac{\pi}{2}]$ represents rotation angle and satisfy the normalization condition
$|\cos(\theta_{i,k})|^2 + |\sin(\theta_{i,k})|^2 = 1$ with

$|cos(\theta_{i,k})|^2$ *the probability of rejecting* k^{th} *gene of the* i^{th} *quantum moth*
$|sin(\theta_{i,k})|^2$*the probability of selecting* k^{th} *gene of the* i^{th} *quantum moth*

Simultaneously, the flame population containing *n* and the number *d* of Q-bits. It is represented as $QF = \{Qf_1, Qf_2, \ldots, Qf_d\}$ where Qf_j (j = 1, 2, ... n) is the j^{th} flame. Each Qf_j quantum flame represents as follows (Eq.3):

$$Qf_i = \begin{vmatrix} \cos(\omega_{j,1}) & \cos(\omega_{j,2}) & \cos(\omega_{j,3}) & \ldots\ldots & \cos(\omega_{j,d}) \\ \sin(\omega_{j,1}) & \sin(\omega_{j,2}) & \sin(\omega_{j,3}) & \ldots\ldots & \sin(\omega_{j,d}) \end{vmatrix}$$

Where, $\omega_{j,k}$ (k = 1, ..., d) satisfy the normalization condition
$|cos(\omega_{i,k})|^2 + |sin(\omega_{i,k})|^2 = 1$ and $\omega_{i,k} \epsilon [0, \frac{\pi}{2}]$

2.2. The Hybrid QMFOA Approach

The basic structure of the QMFOA has presented in this paper is described by Algorithm 1.

3. RESULTS AND DISCUSSIONS

Accuracy is one of the evaluation criteria of the classification model. The accuracy of the classification is the overall correctness of the classifier and is defined as the sum of the true correct cancer classifications divided by the total number of classifications. The accuracy of the classification is calculated according to Eq. 4.

$$Classification\ Accuracy = \frac{CC}{N} \times 100$$

Where, N is the total number of the instances in the initial Microarray dataset and *CC* refers to correct classified instances.

3.1. Dataset

Table 1 presents detailed characteristics of these gene expression datasets in terms of the number of classes, the number of genes, sample size, reference, and a brief description.

3.2. Parameter Settings

The parameters used in QMFOA are displayed in Table 2. The QMFOA used the mMRM as pre-filters to select the 100 top-ranked genes from all. In this study, to perform our experiments, the number of runs is 10 times on each dataset.

3.3. Experimental Results and Analysis

The QMFOA is evaluated on two kinds of benchmark Microarray cancer data, which are binary class and multi class datasets, in order to evaluate the performance and prove the effectiveness of the QMFOA to the gene selection problem. To achieve this, we made a couple of comparisons with some recently published algorithms.

Algorithm 1: QMFOA pseudo-code.

Input:
Data: Dataset; ▷ Data set
dsubset: set of genes; ▷ The set of genes of genes by pre-filter mRMR (d: genes).
Pop_Size: integer; ▷ population size.
QM [*Pop_Size*], *QF* [*Pop_Size*] of Quantum individual;
▷ *QM*: Quantum moth and *QF*: Quantum flame.
BM [*Pop_Size*], *BF* [*Pop_Size*] of Binary individual;
▷ *BM*: Binary moth and *BF*: Binary flame.
FM [*Pop_Size*], *FF* [*Pop_Size*] of Float;
▷Fitness of moth population *FM* and flame population *FF*.

Output:
Subsetbest : set of genes; ▷ Best subset of genes

1: Normalization_Min-Max (Data); ▷Normalization of the dataset
2: *dsubset* ←mRMR (*Data*); ▷ the subset of genes that obtained by pre-filter of the mRMR
3: **for** ($k \leftarrow 1$ **to** *Pop_Size*) **do**
4: *Initialization* ($QM^0_{i,k}, |d_{subset}|$ ▷ Initialize a population QM^0 of $|d_{subset}|$quantum moths
5: **end for**
6: *Iteration* ← 0
7: **repeat**
8: *N br_Flames* is calculated using by [16])
9: **for** ($i \leftarrow 1$ **to** *Pop_Size*) **do**
10: BM_i^t ←Transformation ($QM_i^t, |d_{subset}|$); ▷Make a BM^t of $|d_{subset}|$ from QM^t by Transformation function that defined in Algorithm 2
11: *FM* [i] ←Evaluate BM_i^t to fitness function;
12: **end for**
13: **if** *Iteration* ← 0 **then**
14: *QF; BF; FF* ← *sort* (FM^t, QM^t, BM^t); ▷Sort QM^t, BM^t in ascending order by the fitness function / (t = 0)
15: **else**

16: $QF;\ BF;\ FF \leftarrow sort\ (FM^t;\ [QF;QM^t];\ [BF;BM^t])$; ▷ Sort $[QF;QM^t]$, $[BF;BM^t]$ in ascending order by the fitness function i.e., among moth population ($t-1$, t)
17: **end if**
18: **for** ($i \leftarrow 1$ **to** *Pop_Size*) **do**
19: **for** ($j \leftarrow 1$ **to** *N br_Flames*) **do**
20: $Dist_{i;j} \leftarrow Distance(\ BM_i^t\ ,\ BF_j^t$ Calculate the distance between the i^{th} moth (BM_i^t) , the j^{th} flame (BF_j^t)
21: **end for**
22: $Qm_i^{t+1} \leftarrow Update\_Q\ (Qm_i^t, Dist_{i;j}\ , |d_{subset}|)$ ▷Update the quantum moth by *Update_Q* function that defined in Algorithm 3
23: **end for**
24: $Subset_{best} \leftarrow QF_0, BF_0, FF$ [0]
25: $Iteration \leftarrow Iteration + 1$
26: **until** (*Iteration* >*Max_iteration*)
27: **return** $Subset_{best}$

3.3.1. Results obtained by QMFOA on binary class datasets

In this work, we compare QMFOA with well-known gene selection algorithms published in the literature that have applied to binary datasets, which present in Table 3, like PCC-BPSO and PCC-GA [12], MOBBA_LS [6], GBC [3], ICA-ABC [4], MIM-AGA [15].
Table 3 illustrates the comparison of the experimental results between the QMFOA and other gene selection algorithms that applied to binary datasets, in terms of the best, worst, average and standard deviation (S.D.) of the number of genes selected and the classification accuracy. Cells with unknown values, to our knowledge, are represented with the '-' character.

Algorithm 2: Transformation pseudo-code.
Input:
Qm_i *:* quantum individual; ▷Qm_i: quantum moth.
d: integer; ▷ where, d: the maximum number of genes in a moth

Output:
Bm_i : binary individual; ▷ Bm_i: binary moth.

1: **function** *Transformation* (Qm_i, *d*)
2: **for** ($j \leftarrow 1$ **to** *d*) **do**
3: $threshold \leftarrow random\ (0, 1)$ ▷ Generate a random real value between 0 and 1
4: **if** ($threshold > |Cos\ (\theta_{i,j})|^2$ **then**
5: $Bm_{i;j} \leftarrow 1$;
6: **else**
7: $Bm_{i;j} \leftarrow 0$;
8: **end if**
9: **end for**
10: **return** Bm_i
11: **end function**

Algorithm 3: Updating-Quantum pseudo-code.
Input:
Qm_i^t : quantum moth at generation *t*;
$Dist_{i;j}$: real; ▷ $Dist_{i,j}$: Distance between i^{th} moth and j^{th} flame at generation *t*.
d: integer; ▷ where *d*: the maximum number of genes in the moth $|d_{subset}|$.

Output:
Qm_i^{t+1} : quantum moth of the next generation (t+1);
BM_i^{t+1} : binary moth of the next generation (t+1);

1: **function** UPDATE_Q (Qm_i^t, $Dist_{i;j}$, *d*)
2: **for** ($k \leftarrow 1$ **to** *d*) **do**
3: **if** *fitness* ($Bm_{i;k}$) < *fitness*($Bf_{j;k}$) **then**
4: $\alpha \leftarrow rand\ (0, 2 * k * \pi)$
5: $\theta_{i,k}^{t+1} \leftarrow Dist_{i;j} * \alpha$
6: **else**
7: $\theta_{i,k}^{t+1} \leftarrow \theta_{i,k}^t$
8: **end if**
9: **end for**
10: $Bm_i^{t+1} \leftarrow Transformation\ (Qm_i^{t+1}, d)$ ▷Apply *Transformation* on a new quantum moth Qm_i^{t+1}
11: **return** Bm_i^{t+1}, Qm_i^{t+1}
12: **end function**

As can be seen in Table 3, for Leukemia1, four algorithms (PCC-BPSO, PCC-GA, GBC and our algorithm) can obtain 100% classification accuracy. For MOBBA_LS, ICA-ABC, and MIM-AGA methods selected 3, 5 and 7 genes and achieved 97.1%, 96.43%, and 97.68% classification accuracy, respectively. In contrast, our algorithm selects 32 genes and achieves 100% classification accuracy. In the best-obtained results for Leukemia1, the QMFOA obtained a slightly larger amount of genes than PCC-BPSO and GBC. For Prostate_Tumor, our method has achieved the highest accuracy (100%), but the MOBBA_LS method selected 6 genes and achieved 94.10% classification accuracy, in terms of accuracy, it is very far from QMFOA, which is better than all methods.

In addition, for CNS, the QMFOA method has achieved the highest accuracy that is better than all methods by 100%, at all classification accuracy. For Colon, the ICA-ABC method selected 12 genes and achieved 90.22% average accuracy.

Table 1: Description for the test gene expression datasets.

Dataset Name	Samples	Features	Classes	Notes	Source
CNS	60	7129	2 (Binary class)	'MS': 39, 'TF': 21	[27]
Colon	62	2000	2 (Binary class)	'Tumor': 40, 'Normal': 22	[2]
Leukemia1	72	7129	2 (Binary class)	'ALL': 47, 'AML': 25	[11]
Breast	97	24481	2 (Binary class)	'non-relapse': 51, 'relapse': 46)	[27]
Ovarian	253	15154	2 (Binary class)	'Cancer': 162, 'Normal': 91	[22]
Prostate_Tumor	102	10509	2 (Binary class)	'Normal': 52, 'Tumor': 50	[24]

Table 2: QMFOA parameters for gene subset selection and classification.

parameters	**Setting value**
Population size	50
Normalization interval	[-1,1]
GO_1	0.70
GO_2	0.30
Random angle (Archimedes spiral)	$[0, 6\pi]$ / $k = 3$
Number of generation (iteration)	30
Top-ranked genes by mMRM	100

In contrast, the QMFOA is better than all methods; it selects 30.67 genes and achieves 100% classification accuracy. For Breast, the MIM-AGA method selected 216 genes and achieved 95.21% average accuracy. But, our algorithm selects 27.73 average genes and achieves 81.44% classification accuracy. On the other hand, the QMFOA selected 0.12% of the genes in terms of the number of genes selected by the MIM-AGA.
Finlay, For the Ovarian dataset, the PCC-BPSO, PCC-GA, and QMFOA can provide 100% in terms of best accuracy with the number of genes selected being 17, 22, and 17, respectively. The QMFOA can provide more than 99% for average accuracy and less than 20 selected genes.
Based on the above analysis, in this comparison, we can conclude that QMFOA has given better results than other algorithms in terms of the classification accuracy and the number of genes selected.

Table 3: Comparison of experimental results obtained by MIM-mMFA with other methods for binary class datasets.

Algorithms		Dataset	Leukemia1	Prostate Tumor	CNS	Colon	Breast	Ovarian
QMFOA	Accuracy	Best	100,00	100,00	100,00	100,00	81,44	100,00
		Worst	100,00	98,02	100,00	100,00	74,23	98,42
		Avg.	100,00	99,87	100,00	100,00	77,53	99,37
		S.D.	0,00	0,51	0,00	0,00	2,07	0,44
	# Genes	Best	32,00	26,00	28,00	27,00	22,00	17,00
		Worst	41,00	39,00	40,00	34,00	33,00	24,00
		Avg.	36,47	32,60	31,27	30,67	27,73	20,60
		S.D.	3,00	4,14	3,63	2,09	2,96	2,16
PCC-BPSO	Accuracy	Best	100,00	97,06	98,33	91,94	90,72	100,00
	# Genes	Best	18,00	33,00	39,00	25,00	41,00	17,00
PCC-GA	Accuracy	Best	100,00	96,08	98,33	91,94	88,66	100,00
	# Genes	Best	35,00	26,00	48,00	29,00	38,00	22,00
MOBBA_LS	Accuracy	Best	97,10	94,10	-	-	-	-
	# Genes	Best	3,00	6,00	-	-	-	-
GBC	Accuracy	Best	100,00	-	-	98,38	-	-
		worst	93,05	-	-	91,93	-	-
		Avg	96,43	-	-	94,62	-	-
	# Genes	Best	5,00	-	-	20,00	-	-
ICA-ABC	Accuracy	Best	98,21	97,88	-	97,34	-	-
		worst	55,76	77,81	-	82,34	-	-
		Avg	83,22	82,34	-	90,22	-	-
	# Genes	Best	12,00	20,00	-	12,00	-	-
MIM-AGA	Accuracy	Best	97,68	97,69	-	89,09	95,21	-
	# Genes	Best	7,00	93,00	-	19,00	216,00	-

4. CONCLUSIONS

In this work, we have presented a new hybrid technique between quantum computing and the Moth flam optimization called quantum Moth flam optimization algorithm (QMFOA) for gene selection and classification of high dimensional datasets. Therefore, the goal of this work is to provide a new bio-inspired algorithm to solve gene selection problems.

The QMFOA consists of two stages. In the first stage, we used the mRMR as a pre-filter method to rank the gene scores and select the 100 top genes as inputs of the second stage. The overall objective of this work is to select a smaller number of genes and obtain a classification accuracy similar to or better than that obtained by using all genes.

The experimental results of QMFOA on six binary class datasets have shown that our algorithm can find useful informative genes than all other compared algorithms in terms of classification accuracy, i.e. is better than all other compared algorithms, and also able to deliver competitive results in terms of the number of genes.

REFERENCES

[1] Abderrahim, A., Talbi, E.G., Khaled, M., 2012. Hybridization of genetic and quantum algorithm for gene selection and classification of microarray data. International Journal of Foundations of Computer Science 23, 431–444.

[2] Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., Levine, A.J., 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proceedings of the National Academy of Sciences 96, 6745–6750.

[3] Alshamlan, H.M., Badr, G.H., Alohali, Y.A., 2015. Genetic bee colony (gbc) algorithm: A new gene selection method for microarray cancer classification. Computational biology and chemistry 56, 49–60.

[4] Aziz, R., Verma, C., Jha, M., Srivastava, N., 2017. Artificial neural network classification of microarray data using new hybrid gene selection method. International Journal of Data Mining and Bioinformatics 17, 42–65.

[5] Cover, T.M., Van Campenhout, J.M., 1977. On the possible orderings in the measurement selection problem. IEEE transactions on systems, man, and cybernetics 7, 657–661.

[6] Dashtban, M., Balafar, M., Suravajhala, P., 2018. Gene selection for tumor classification using a novel bio-inspired multi-objective approach. Genomics 110, 10–17.

[7] Davies, S., Russell, S., 1994. Np-completeness of searches for smallest possible feature sets, in: AAAI Symposium on Intelligent Relevance, AAAI Press. pp. 37–39.

[8] Deng, X., Li, Y., Weng, J., Zhang, J., 2019. Feature selection for text classification: A review. Multimedia Tools and Applications 78, 3797–3816.

[9] Du, D., Li, K., Li, X., Fei, M., 2014. A novel forward gene selection algorithm for microarray data. Neurocomputing 133, 446–458.

[10] Duval, B., Hao, J.K., Hernandez Hernandez, J.C., 2009. A memetic algorithm for gene selection and molecular classification of cancer, in: Proceedings of the 11th Annual conference on Genetic and evolutionary computation, ACM. pp. 201–208.

[11] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al., 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. science 286, 531–537.

[12] Hameed, S.S., Muhammad, F.F., Hassan, R., Saeed, F., 2018. Gene selection and classification in microarray datasets using a hybrid approach of pcc-bpso/ga with multi classifiers. JCS 14, 868–880.

[13] Han, K.H., Kim, J.H., 2002. Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. IEEE transactions on evolutionary computation 6, 580–593.

[14] Jain, A., Nandakumar, K., Ross, A., 2005. Score normalization in multimodal biometric systems. Pattern recognition 38, 2270–2285.

[15] Lu, H., Chen, J., Yan, K., Jin, Q., Xue, Y., Gao, Z., 2017. A hybrid feature selection algorithm for gene expression data classification.

[16] Mirjalili, S., 2015. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. Knowledge-Based Systems 89, 228–249.

[17] Mudaliar, P.U., Patil, T.A., Thete, S.S., Moholkar, K.P., 2015. A fast clustering based feature subset selection algorithm for high dimensional data. International journal of emerging trend in engineering and basic science 2, 494–499.

[18] Mundra, P.A., Rajapakse, J.C., 2010. Gene and sample selection for cancer classification with support vectors based t-statistic. Neurocomputing 73, 2353–2362.

[19] Narayanan, A., 1999. Quantum computing for beginners, in: Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406), IEEE. pp. 2231–2238.

[20] Nielsen, M.A., Chuang, I., 2002. Quantum computation and quantum information.

[21] Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minredundancy. IEEE Transactions on Pattern Analysis & Machine Intelligence , 1226–1238.

[22] Petricoin III, E.F., Ardekani, A.M., Hitt, B.A., Levine, P.J., Fusaro, V.A., Steinberg, S.M., Mills, G.B., Simone, C., Fishman, D.A., Kohn, E.C., et al., 2002. Use of proteomic patterns in serum to identify ovarian cancer. The lancet 359, 572–577

[23] Sardana, M., Agrawal, R., Kaur, B., 2016. A hybrid of clustering and quantum genetic algorithm for relevant genes selection for cancer microarray data. International Journal of Knowledge-based and Intelligent Engineering Systems 20, 161–173.

[24] Singh, D., Febbo, P.G., Ross, K., Jackson, D.G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A.A., D'Amico, A.V., Richie, J.P., et al., 2002. Gene expression correlates of clinical prostate cancer behavior. Cancer cell 1, 203–209.

[25] Wang, H., Niu, B., 2017. A novel bacterial algorithm with randomness control for feature selection in classification. Neurocomputing 228, 176–186.

[26] Xi, M., Sun, J., Liu, L., Fan, F., Wu, X., 2016. Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine. Computational and mathematical Methods in Medicine 2016.

[27] Zhu, Z., Ong, Y.S., Dash, M., 2007. Markov blanket-embedded genetic algorithm for gene selection. Pattern Recognition 40, 3236–3248.

Who Needs an Encryption Backdoor: Why Americans want Security over Privacy.

Robert E. Endeley
Adjunct Faculty, Dunwoody College of Technology
rendeley@dunwoody.edu

Abstract - **A qualitative analysis study that examined the views and opinions of non-technology professionals in the U.S. regarding government and law enforcement agencies' demand for legislation that will allow them to snoop on online private communications of smartphone users. Governments would prefer exclusive access to encryption technologies, called a backdoor, to use in accessing messages. Technology professionals have, however, argued against a backdoor; they claim a backdoor would not only be an infringement of their privacy but that hackers could also take advantage of it. In light of this security and privacy conflict between technology professionals and government's need to access messages in order to thwart potential terror attacks, this study presents the views and opinions of non-technology professionals in the U.S. who are the largest group of smartphone users, on the ensuing encryption debate. Using qualitative descriptive design methodology, a survey of 26 participants was conducted and data was analyzed using Braun and Clarke's six-step process of inductive thematic analysis. Results from this research study showed that non-technology professionals are willing to allow the government to infringe on their privacy if that will guarantee them security.**

Keywords: ***instant messaging, WhatsApp, end-to-end encryption, privacy, government***

I INTRODUCTION

Since smartphones became popular, many instant messaging (IM) services have been launched (Yeboah & Ewur, 2014). Some governments have become concerned about the ubiquity of IM services on mobile phones and their use of end-to-end encryption (E2EE) in safeguarding users' privacy, as it makes eavesdropping harder for them (Endeley, 2018; Michalas, 2017). E2EE ensures messages between communicating parties are secure, free from snooping, and hard to crack (Brantly, 2017). E2EE offers peace of mind to end users as it secures their data in transit and from third parties (Endeley, 2018). The service provider cannot access the messages, which can only be decrypted by the intended recipient (Michalas, 2017; "WhatsApp," 2017).

Governments would prefer special access to encryption technologies, called a backdoor, to use in accessing messages (Michalas, 2017). According to McCarthy (2016), a backdoor is an intentionally engineered gateway into the encryption system to provide an alternative means of accessing the encrypted content. An encryption backdoor may allow third parties to gain access to unencrypted data using certain keys (Abelson et al., 2015). The same backdoor used by an authorized third party such as a government agency authorized by court order may also be vulnerable to an unauthorized attacker who should not have access to the data (Abelson et al., 2015). Governments have emphasized they will only use the backdoor if there is a credible threat to national security (Brantly, 2017). In opposition to governments' proposals for a backdoor, Kern (2012) argued the promise of privacy guaranteed by modern encryption techniques, is, to a great extent, what has expounded the broad use of the internet. Kern further stated common online practices, such as online shopping, banking, and remote terminal services, would largely be impossible without the guarantee of the privacy and confidentiality provided by encryption.

According to Max (2016), governments and security agencies are wrong due to their unfounded belief that strong encryption that protects information on the internet, can at the same time be made weak in order to grant the government access to information.

The encryption and privacy debate heated up more recently following the indictment and conviction of President Donald Trump's former campaign manager, Paul Manafort: United States v. Manafort, District Court, District of Columbia (Novak, 2018). The federal prosecutors accused him amongst other things of witness tampering using the end-to-end encrypted messaging applications WhatsApp and Telegram (Novak, 2018). While E2EE ensures integrity, security, and privacy, it removes opportunities for government surveillance and the capacity to keep the nation secure by intercepting terrorist communications (Rastogi & Hendler, 2017).

According to McCarthy (2016), the Federal Bureau of Investigation (FBI) has been voicing concern that due to barriers such as strong encryption, government's security apparatus has been going dark in its attempt to monitor certain electronic communications and suspected terrorists. McCarthy revealed an increasing awareness of data-related privacy concerns in the aftermath of the Edward Snowden revelations made from 2013 onwards. These revelations purported to show the wide-reaching extent of bulk government surveillance by the U.S. and U.K. security agencies (McCarthy, 2016). McCarthy further stated the world's leading internet communication services providers such as Apple, Google, Facebook, WhatsApp, and Blackberry, have rushed to announce a renewed commitment to customer privacy. These companies all announced plans to implement E2EE on a default basis.

Law enforcement has been advocating for a backdoor into E2EE in IM services, thus undermining privacy and security (McCarthy, 2016). The New York County District Attorney, Cyrus Vance, in a written testimony to the U.S. Senate Judiciary Committee said Apple and Google smartphones should be configured to allow data on these devices to be accessed by law enforcement when it has judicial authorization to do so ("U.S. Department of Homeland Security," 201

Law enforcement agencies such as the FBI have argued to the U.S. Congress that the only way to compel smartphone manufacturers to comply with their request for a backdoor will be through legislation (Barr, 2016).

Much of the literature regarding the effects of E2EE on society has centered on the points of view of cryptographers and law enforcement agencies (Brantly, 2017). An in-depth review of the literature on this debate, however, showed no study had been done before in the U.S. to seek the opinions and views of non-technology professionals. Brantly, 2017 stated the former NSA and CIA Director, General (Ret.) Michael Hayden said, "we will only go as far as the American people allow us, but we will go all the way to that line" (p. 29). General (Ret.) Michael Hayden did not give any details following his statement on the view of the American people regarding encryption backdoors; he left it to anyone's imagination (Brantly, 2017). According to the "U.S. Census Bureau" (2016), technology professionals represent only 2.9 percent of the U.S. labor force.

Non- technology professionals, therefore, represent the largest segment of the labor force, and by inference the largest group of smartphone users ("U.S. Department of Labor," 2019; "U.S. Census Bureau," 2016). There are approximately 152 million working non-

technology professionals in the U.S. who will be impacted and are not aware ("U.S. Department of Labor", 2019; "U.S. Census Bureau", 2016). This study sought to understand the lengths at which nontechnology professionals would want the government to go regarding reading their private messages as a tradeoff for more security.

Studies have shown that non-technology professionals do not understand the impact of creating backdoors into encryption technologies (Sagers, Hosack, & Rowley, 2015; Wei et al., 2016

Non-technology professionals may not think of encryption very much, but it is fundamental to all our lives. Almost everything we do today on the internet uses a secret code, including internet banking, or logging on to Twitter or Facebook; encryption protects all such information. While E2EE protects users' IM from eavesdropping by third parties, full-disk encryption protects data such as photos, texts, emails, contacts, and bank account information from access by rogue individuals who may either steal your device or lay hands on a lost one (Herzberg & Leibowitz, 2016).

Vaziripour et al. (2018) asserted that non-technology professionals lack understanding of what an encrypted chat means and does to guarantee security. This study was, therefore, posited on the central question of whether non- technology professionals understood the impact of government-mandated backdoors on encrypted public messaging services. This study used a qualitative analysis methodology. A qualitative descriptive design was the selected design methodology for this study. A qualitative descriptive design study enabled accurate depictions of participants' views on the impact of government-mandated encryption backdoors (Dews-Farrar, 2018). Additionally, the researcher sought to augment the sparse number of scholarly qualitative descriptive studies concerning end-to- end encryption (E2EE), backdoors, and privacy.

The study intends to explore the following questions which serve as the primary focal points of this study:

- RQ1: Do non-technology professionals in the U.S. understand the impact of creating backdoors into end-to-end encrypted technologies?
- RQ2: What are the perspectives of non-technology professional users of IM applications regarding the argument security comes at a price, namely at the expense of privacy?
- RQ3: To what extent does the knowledge of encryption as a technology in safeguarding consumer privacy affect the use of the internet by non-technology professionals in the U.S.?

II RELATED WORK

Several authors have analyzed the intensifying debate on the proliferation of robust encryption technologies on mobile devices across the globe. In the FBI v. Apple case of 2016, the FBI wanted Apple to rewrite its operating system software (iOS), to disable encryption security features so the FBI could access the data (Elmer-Dewitt, 2016). A Pew survey showed in December 2015, the public sided with the FBI initially, with around 51% arguing Apple should help the FBI unlock the phone, 38% supporting Apple, and 11% not knowing enough about the dispute to form an opinion (Elmer-Dewitt, 2016). However, later polls in February 2016, with diverse methodologies, showed the public sided with Apple (Elmer-Dewitt, 2016). This demonstrated that by Apple making a strong public case in protecting the privacy of its users through the use of encryption, it also educated its user-base on their role in preserving user-privacy (Elmer- Dewitt, 2016). Apple's vigorous defense of its software shifted public opinion to its favor (Elmer-Dewitt, 2016)

According to a report published by the "U.S. Department of Homeland Security" (2015), the U.S. Senate held a hearing on whether recent technological changes have upset the balance between public safety and privacy. Law enforcement officials are becoming increasingly concerned that even after obtaining a warrant from a judge to search for evidence of a crime, they lack the technical means to do so ("U.S. Department of Homeland Security," 2015). This is due to companies increasingly choosing to encrypt devices in such a way the companies themselves are unable to unlock them, even when presented with a valid search warrant ("U.S. Department of Homeland Security," 2015). First, law enforcement agencies are reporting a decreasing ability to intercept real-time communications such as phone calls, text, email, and other types of data-in-transit ("U.S. Department of Homeland Security," 2015). Second, they relate a similar concern about their inability to execute search warrants on encrypted phones, laptops and other devices which contain data-at-rest ("U.S. Department of Homeland Security," 2015). Given this technological evolution, there is a potential impact on the fair and impartial application of the laws to everyone, as certain people are effectively placed outside the law ("U.S. Department of Homeland Security," 2015). The "U.S.

Department of Homeland Security" report concluded mandated technological weaknesses in encryption as proposed by some law enforcement agencies as a means of solving the problem of having exclusive access to these encrypted devices, are both futile and counterproductive. The report concluded Congress was open to reviewing ways to provide law enforcement with judicially-sanctioned access to these platforms without compromising overall security ("U.S. Department of Homeland Security," 2015).

In an article published by WIRED magazine in April 2018, former Microsoft Chief Software Architect and creator of Lotus Notes, Ray Ozzie, made a proposal at Columbia University on how to solve the impasse over secure backdoors between technology companies and law enforcement agencies (Levy, 2018). In his idea named CLEAR, Ozzie stated that his scheme would give law enforcement agencies access to encrypted data without significantly elevating the risks for billions of people who use encrypted devices such as mobile phones (Levy, 2018). Ozzie added the scheme works by technology companies such as Google or Apple generating two complementary keys: one called the vendor's public key, stored in every Android phone or tablet, and the other is called the vendor's private key (Levy, 2018). The private key is stored with Google and protected with the same tamper-proof care Google uses to certify its operating system updates (Levy, 2018). A combination of the private and public key pair can be used to encrypt and decrypt a secret PIN which each user's device automatically generates upon activation. It should be noted Ozzie's scheme attempts to solve the impasse with stored data (data at rest) and not the interception of real-time communications (data in transit) (Levy, 2018). According to Abelson et al. (2015), if law enforcement wants to assure itself access to real-time communications with backdoors, then intruders will also have an easier time getting access to real-time data.

Schneier et al., (2016) have pointed out countries such as the U.S., the U.K., and France seem very interested in mandating backdoors. The impetus to mandate backdoors in encryption products for the countries mentioned above is coming from law enforcement. Security researchers, according to Schneier et al. have, however, argued backdoors are impossible to implement securely and will result in reduced security for everyone. A practical limitation to mandating backdoors as a way of reducing crime is because encryption products come from different parts of the world (Schneier et al., 2016). Anyone attempting to evade encryption backdoors in the U.S., the U.K., or France has a wide variety of foreign encryption products to pick from which can encrypt hard drives, voice and text conversations, virtual private networks (VPN) links and everything else (Schneier et al., 2016). Schneier et al. identified 865 hardware or software encryption

products from 55 countries: 546 of these products, or two-thirds, were from outside the United States. Schneier et al. outlined that most common non-U.S. countries for encryption products were Germany, followed respectively by the United Kingdom, Canada, France, and Sweden. Germany and the Netherlands have publicly disavowed backdoors in all their encryption products.

III METHODOLOGY

The researcher conducted an Institutional Review Board (IRB) approved web and paper-based survey of non- technology professionals in the U.S. Since the study was designed for non-technology professionals, it was possible that not all of the participants will have access to the internet or own a computer. Hence, the need for a paper-based version of the survey. This study used a qualitative analysis methodology. The rationale for selecting a qualitative analysis for this research was based on the diagnosis of the purpose statement. Qualitative descriptive design method was the most appropriate for this research because it sought to gain insight into the views and opinions of non- technology professionals regarding their privacy on public communication platforms. Such an approach was especially useful for researchers wanting to know the "what" and "how" of events (Dews-Farrar, 2018).

A. Research Method and Design Appropriateness

The general population of the study were adult users of mobile phones located in the U.S. and running the latest version of end-to-end encrypted IM service, WhatsApp, on their mobile phones. The IM application WhatsApp was selected for this study because according to Sutikno, Handayani, Stiawan, Riyadi, and Subroto (2016), it is amongst the most favored IM applications endowed with E2EE. Jisha and Jebakumar (2014) stated WhatsApp is the fastest-growing IM application as most young people are moving away from Facebook. WhatsApp enjoys global favorability with a user base of over 1.5 billion subscribers. It is also the first application ever to implement E2EE to this scale (Rastogi & Hendler, 2017). From the target population, the researcher chose a research sample of 26 participants who met the criteria for participation.

B. Population, Sampling, and Data Collection Procedures and Rationale

This study posited non-technology professionals have a limited understanding of the consequences of a government- mandated backdoor into encryption technologies. The sample size of 26 for this study met the saturation limit in qualitative descriptive research as shown in similar research carried out by Dews-Farrar (2018) using the same design. The participants had to meet the following criteria: (a) Participants had to be owners of a mobile phone running the latest version of the WhatsApp application. (b) They had to be willing to participate in an online survey, or a face-to- face interview with the researcher. (c) They had to be non-technology professionals who at the time of this study did not have any experience working in technology or hold any diploma or certification in computer science, computer security or computer networking. (d) Participants had to be willing to give honest accounts of their views and opinions about privacy and national security.

C. Sampling

After obtaining IRB approval, the researcher initiated the participant recruitment process. Purposeful sampling, specifically the snowball sampling methodology, was used for the selection of the 26 participants. Purposeful sampling involves the selection of individuals who are qualified to provide in-depth information about the phenomenon being researched. Snowball sampling is a sub-category of purposeful sampling and has the advantage that after observing the initial participants, the researcher asks for assistance from the participants to help identify people with a similar trait (Creswell, 2015). Non-technology professionals like themselves who meet the requirements for the sample population (Creswell, 2015).

Snowball sampling is a non-probability sampling methodology. Rashidi, Vaniea, and Camp (2016) used snowball sampling in a study on privacy setting usage in WhatsApp application to recruit participants. The study by Rashidi et al. yielded relevant results which have contributed to the body of literature on how users manage privacy settings on IM applications. Snowball sampling generally consists of two steps:

1. The researcher identifies the potential participants in the population. Often, only a handful.
2. The researcher asks the identified participants to recruit other participants. The chain continues until the sample size is reached.

According to the Bureau of Labor Statistics of the "U.S. Department of Labor" (2017), States or jurisdictions with the highest location quotient for information technology experts are Virginia with 4.71, Maryland with 2.50, and the District of Columbia with 2.25. The location quotient is a way of quantifying how concentrated a particular industry or occupation is in a particular region or State, in reference to the entire nation. The States ranking with the lowest location quotient for information security experts are New Mexico, Missouri, and Colorado ranking 1.33, 1.37, and 1.41 respectively. The State of Minnesota falls in the middle of the rankings with 1.65 as its location quotient for information security experts. The average rankings in the distribution of technology professionals in Minnesota in relation to the rest of the country make it an ideal candidate for the target population of this study

D. Limitations

There were two limitations related to the data analysis of this study.

1. The sample population of this study was limited to a single state, Minnesota. The interpretation of results is affected by this limitation because it is not known if location quotient alone or the large population of healthcare workers in Minnesota introduced biases in the results.
2. Snowball sampling is a useful sampling methodology when it is not possible to use more traditional survey techniques. Snowball sampling technique, however, has its limitations. According to Sharma (2017), since snowball sampling does not select units for inclusion in the sample based on random selection, unlike the probability sampling technique, it is impossible to determine the possible sampling error and make generalizations from the sample to the population.

IV RESULTS

B. Pilot Study

A pilot study was carried out to establish the comprehensibility, validity, and reliability of the survey questions. The pilot study for this research consisted of five preliminary participants. The informed consent agreement was given to each participant, and the researcher obtained approval from all five participants. The pilot study revealed that the participants would be better served by defining key terms such as encryption, end-to-end encryption, and backdoors in the participant consent form before they got to the survey questions. Feedback from the pilot study was used to revise the final wording in the survey questions. The findings of this pilot study demonstrated the functionality of the survey instrument and the interest in the research by the target population. Thus, after adjustments to the survey from recommendations of the pilot study participants, the researcher concluded that the survey instrument was valid for this study's topic and served to answer the specific problems posited.

C. *Findings*

Forty-six participants in total responded to the survey between May 1, 2019, and May 10, 2019. Twenty of the respondents to the survey were eliminated during the data analysis phase because they did not meet the survey criteria. Twenty-six participants completed all the survey questions. Only responses of survey participants who met the survey criteria and completed all the questions are included in this article. The online web tool SurveyMonkey was used for preliminary analysis before respondent data were imported into NVivo data analysis software for full analysis.

D. *Demographics*

Out of the 26 participants who completed the survey, 12 (46.15%) were female, while 14 (53.85%) were male. Participants ranged in age from 25 – 70 years. The largest group of participants were between the ages of 45 - 54 years (46.15%), with the smallest between the ages 24 – 35 (7.69%) and 65 – 74 (7.69%). Figure 1 below shows the age distribution of participants. All 26 participants (100%) resided in the state of Minnesota, held no degree or certification in computer science, and their jobs did not include information technology-related activities. Table 1 below also displays a demographic summary of the participants.

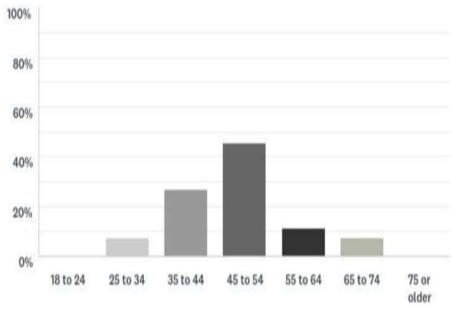

Fig 1. A bar chart showing the age distribution of participants

TABLE 1. DEMOGRAPHIC SUMMARY OF STUDY

Participant	Reside in Minnesota?	Hold computer certification?	Does your job include information technology related activities?	Age Group	Gende
P1	Yes	No	No	55 to 64	Male
P2	Yes	No	No	55 to 64	Femal
P3	Yes	No	No	35 to 44	Male
P4	Yes	No	No	35 to 44	Male
P5	Yes	No	No	45 to 54	Male
P6	Yes	No	No	45 to 54	Male
P7	Yes	No	No	45 to 54	Male
P8	Yes	No	No	45 to 54	Male
P9	Yes	No	No	45 to 54	Femal
P10	Yes	No	No	55 to 64	Male
P11	Yes	No	No	45 to 54	Male
P12	Yes	No	No	35 to 44	Male
P13	Yes	No	No	45 to 54	Male
P14	Yes	No	No	25 to 34	Femal
P15	Yes	No	No	45 to 54	Femal
P16	Yes	No	No	65 to 74	Femal
P17	Yes	No	No	45 to 54	Male
P18	Yes	No	No	35 to 44	Femal
P19	Yes	No	No	35 to 44	Femal
P20	Yes	No	No	45 to 54	Femal
P21	Yes	No	No	45 to 54	Femal
P22	Yes	No	No	65 to 74	Male
P23	Yes	No	No	35 to 44	Femal
P24	Yes	No	No	45 to 54	Male
P25	Yes	No	No	35 to 44	Femal
P26	Yes	No	No	25 to 34	Femal

Braun and Clarke's six-step inductive thematic data analysis approach was used to analyze the data. NVivo 12 Plus data analysis software for Windows was used in coding and analyzing the data for this study. Six themes emerged from the data analysis, namely: government and privacy, information, encryption, activities, communications, and social media. The themes were representative of participant- generated conceptualizations regarding the phenomenon encryption, backdoors, and privacy. Figure 2 below gives a visual representation of the six major themes generated from

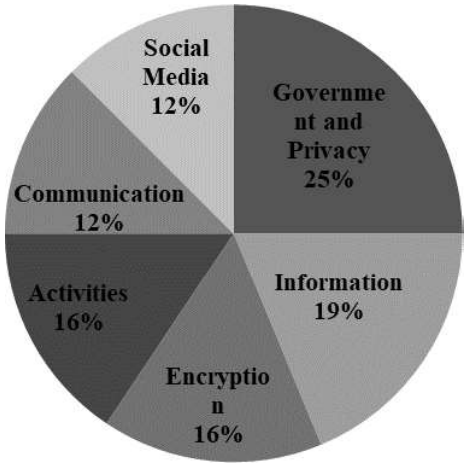

Fig. 2. A pie chart of the six themes generated from the study

the refinement of the initial codes through the process of eliminating redundancies and analysis (Creswell, 2015).

Some candidate themes were merged to form more coherent and meaningful themes; government and privacy themes were merged to form a single theme. The themes were validated as having a connection to the research questions and the overall research problem. The finalization of phase four led to phase five, which was to refine the themes and present them for data analysis.

Braun and Carke (2006) asserted that the researcher should not only be able to explain the relationship between the themes and the research questions but should additionally be prepared to construct an analysis of each theme. Each theme should portray the participants' collective perceptions of the phenomenon under research Braun and Clarke (2006). Figure 3 below shows the relationship between the research questions, codes, and themes.

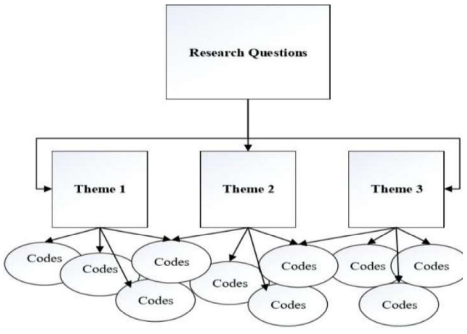

Fig. 3. The relationship between research questions, codes, and themes

E. Findings and Interpretations

Six themes emerged from the data analysis represent the survey participants' conceptualization regarding the phenomenon of the government exploring ways to implement encryption backdoors in popular messaging applications such as WhatsApp.

The themes provided closure for the three research questions that were the basis for this study. Data analysis of survey answers to RQ1 produced a single theme, government, and privacy.

1) Theme: Government and Privacy.

Twenty-three participants (88%) who coded for these themes showed a clear understanding of what an encryption backdoor meant and its impact on their privacy. This was significant to this research because it answered RQ1. Eleven of the participants (42%) were not opposed to the government adding a backdoor to read their private messages in order to keep them safe, especially if there is a credible threat against public safety. Further, these findings are even more significant because they validate the purpose of the research, which was to raise awareness for non- technology professional users of mobile devices on the benefits of encryption to privacy. When participants were asked at the end of the survey if this study had increased their knowledge or awareness of the benefits of end-to-end encryption to your privacy, 23 of the 26 of the participants (89%) responded that it had increased by a moderate amount, a lot, or a great deal. See table 3 below.

TABLE 2. HOW MUCH THE SURVEY IMPACTED PARTICIPANT KNOWLEDGE ON E2EE AND PRIVACY

Answer Choices	Responses	
A great deal	46.15%	12
A lot	23.08%	6
A moderate amount	19.23%	5
A little	11.54%	3
None at all	0.00%	0
Total		**26**

2) Theme: Information.

The sentiments expressed by 18 participants (69%) on this theme was neutral or mixed. Participants said that while they would like their private messages to remain private, they also do not mind the government stepping in to their private information in order to keep the country safe. This theme answered RQ2. Participants asserted through their responses that they are willing to allow the government to infringe on their privacy if that will guarantee them safety.

3) Theme: Activities.

Participants all expressed comfort in letting their security be of a higher priority than their privacy; thus, endorsing the government's intent to monitor electronic activities. This theme also answered RQ2.

4) Theme: Communications.

There were four participants (15%) who coded for this theme. Unlike participants who coded for the activities theme by expressing their comfort with government surveillance in exchange for security, communications participants were decisively against giving up their private communications in exchange for more security. This theme also answered RQ2.

5) Theme: Social Media.

Four participants (15%) expressed their views and opinions on how their knowledge of encryption will affect their use of social media applications such as WhatsApp.

This theme on social media was in response to RQ3, which asked participants to what extent the knowledge of encryption as a technology in safeguarding consumer privacy affect their use of the internet. Participants expressed more confidence in the use of the internet, knowing that encryption helps protect their communications.

6) Theme: Encryption.

Five participants (19%) coded for encryption. This theme was also in response to RQ3. Sixty percent of the participants who coded for this theme were concerned that terrorists could master encryption technology and use it to cause harm to society. In addressing RQ3, participants all agreed their knowledge of encryption would affect their use of the internet by increasing the confidence they have in the privacy of online communications. In addition to the two themes that emerged regarding RQ3, participants were given a layman's definition of encryption technology in the survey and asked if they were aware that popular websites such as Twitter, Facebook, or even their banking operations are all protected from hackers by encryption. Twenty-one participants (81%) answered yes, while five participants (19%) answered no. See figure 4 below.

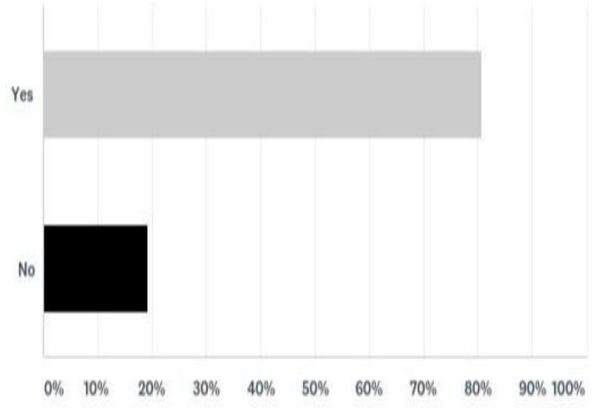

Fig. 4. A distribution graph showing participants knowledge on whether or not they knew encryption was used in protecting their data on popular websites such as Facebook and Twitter.

V. CONCLUSION

This study was relevant not only because it was aimed at filling some of the gaps in the literature regarding the opinions and views of non-technology professionals on the effects of end-to-end encryption (E2EE) on society but, it also confirmed and challenged previous studies on some of the privacy concerns expressed by U.S. mobile device users. Open-ended questions provided participants a means to best voice their experiences unrestricted by the influence of the researcher or past research findings (Creswell, 2015).

The results of this research study have confirmed some of the privacy concerns expressed by mobile device users, as mentioned by Rastogi and Handler (2017) and Elmer-Dewitt (2016). According to Elmer-Dewitt, following the standoff between Apple v.

FBI over access to the iPhone of the San Bernardino shooter, Americans, by a small margin (46% to 35%) support the government's right to access data in smartphones in order to protect the country against terror threats. This research study has also demonstrated that while concerned with their privacy, non-technology professionals are willing to allow the government to access their private messages if they have to do so in order to preserve national security.

Vaziripour et al. (2018) asserted the lack of understanding by non-technology professionals of what an encrypted chat means, as the reason for the none adoption of E2EE; this study proved the contrary. Eighty-one percent of participants who completed this research study said they were aware of what encryption was, and that is was used on most popular websites such as Twitter and Facebook to safeguard their private and sensitive information

Twenty-three participants (88%) showed a clear understanding of what an encryption backdoor meant and its impact on their privacy. A majority of participants (42%) were also not opposed to the government adding a backdoor to read their private messages in order to keep them safe, especially if there is a credible threat against public safety. Government has maintained that they will only use this method of access if there is a credible threat to public safety (Brantly, 2017).

Also, a majority participants (69%) said that while they would like their private messages to remain private, they also do not mind the government stepping in to their private information in order to keep the country safe.

A. Implications and Findings

The results of this study may also help educate the everyday user of the internet on the benefits of E2EE in their daily communications on mobile devices. Participants of this research study have said that it has significantly increased their knowledge of encryption. This creation of awareness and expectation of privacy guaranteed by strong encryption for the everyday user of the internet may also drive more technology companies to adopt E2EE, as was the case after the Edward Snowden leaks in 2013 (McCarthy, 2016). Another benefit this study may bring is, non-technology professionals may increase their adoption of using the internet for personal transactions such as paying bills, online banking, and money transfers once they are aware and understand the benefits of strong encryption on the internet. Participants of this study have expressed an increase in confidence in their privacy on the internet, knowing that encryption guarantees such privacy.

B. Strengths of this Study

There were four strengths in this research study. The first strength was that the researcher achieved data saturation through the participation of 26 working non- technology professionals who participated in the survey. The researcher's employment of structured and open-ended survey questions, a typical approach in qualitative inquiry, yielded detailed and insightful portrayals of the participants lived experiences and generated substantial data.

The second strength involved the use of a pilot study. According to Abdul, Othman, Mohamad, Lim, and Yusof (2017), survey questions could be strengthened by piloting the surveys. It can also help identify if there are flaws or limitations within the survey design that allow necessary modifications to the major study. (Abdul et al., 2017). The pilot study for this research established the comprehensibility, validity, and reliability of the survey questions. The survey questions were revisited to allow quality data and more in-depth responses from the participants.The third strength of this research study was the utilization of manual coding and subsequently, NVivo 12 Plus data analysis software. Prior to utilizing the NVivo 12 Plus qualitative data analysis program, each survey submission was read several times, portions of the text were highlighted in the Microsoft Word documents, and preliminary codes were identified in the right margin of the transcripts. Qualitative analysis depends to a good extent on the subjective interpretations of the researcher. Therefore, a combination of personal judgment and software was used to bring objectivity to the coding process. The utilization of NVivo data software increased the reliability and validity of the study.

The fourth strength was the large amount of qualitative data collected and the detailed descriptions of the participants' recounts and subsequent themes. The data allowed for thorough mining of codes during data analysis and subsequent validation with NVivo data analysis software. The research study was also able to capture participant's knowledge and awareness of E2EE at the beginning of the survey, and also evaluate if they have gained any additional knowledge during the course of the survey at the end.

C. Recommendations

This research study intends to augment the limited number of qualitative descriptive studies regarding the opinions and views of non-technology professionals on the benefits of E2EE on society. The results from this study have revealed insightful accounts of 26 non-technology professionals in the U.S. on E2EE, backdoors, and privacy. Based on the results of this research study the following are recommendations for future research:

1) Extend the research sample area beyond the state of Minnesota to other states. The State of Minnesota falls in the middle of the rankings for the location quotient for information security experts ("U.S. Department of Labor," 2017). Therefore, it is recommended that states with the highest location quotient for information security experts such as Virginia and Maryland, as well as states with the lowest location quotient for information security experts such as New Mexico, Missouri, and Colorado, be sampled.
2) Perform a quantitative study of non-technology professionals with a random sample of participants distributed across the US. This would eliminate some of the inherent weaknesses expressed in the limitations of this research study on the snowball sampling methodology.
3) Include the influence of gender, age, ethnicity, or level of education on participants' views on the government's demand for a backdoor into encryption systems could be an interesting area of research. Similar studies carried out in Iran by Vaziripour et al. (2018) on the Telegram IM application showed that skewed demographics might have influenced the results of the research.
4) Expand this research study internationally, into other countries with less cellphone penetration than the U.S. As mentioned by Schneier et al. (2016), encryption is now a global phenomenon. Laws in the U.S. mandating backdoors into encryption systems will primarily affect only U.S. users of encryption products made in the U.S. (Schneier et al., 2016). Smartphone users in other countries rely on other products. The literature review conducted for this research study found out countries such as Germany, United Kingdom, Canada, France, and Sweden also produce a lot of encryption products (Schneier et al., 2016). Researching the perspectives of people of other countries on this topic would certainly add value to the body of literature in cybersecurity.

REFERENCES

Abelson, H., Anderson, R., Bellovin, S. M., Benalo, J., Blaze, M., Diffie, W.,… Weitzner, D. J. (2015). Keys under doormats: Mandating insecurity by requiring government access to all data and communications. *Computer Science and Artificial Intelligence Laboratory Technical Report*, MIT-CSAIL-TR-2015-026. doi: http://hdl.handle.net/1721.1/97690

Abdul, M. M, Othman, M., Mohamad, S. F., Lim, S. A., & Yusof, A. (2017). Piloting for interviews in qualitative research: Operationalization and Lessons Learnt. International Journal of Academic Research in Business and Social Sciences. 7(*4*). doi:10.6007/IJARBSS/v7-i4/2916

Barr, A. C. (2016). Guardians of Your Galaxy S7: Encryption backdoors and the first amendment. *Minn.L. Rev.*, *101*, 301-383. Retrieved from http://www.minnesotalawreview.org/wp-content/uploads/2016/11/Barr.pdf

Brantly, A. F. (2017, August). Banning encryption to stop terrorists: A worse than futile exercise. *Combating Terrorism Center at West Point, CTCSENTINEL, 10*(7), 29-35. Retrieved from https://ctc.usma.edu/banning-encryption- to-stop-terrorists-a-worse-than-futile-exercise/

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

Creswell, J. W. (2015). *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Education, Inc.

Dews-Farrar, V. (2018). *Students' reflections and experiences in online learning: A qualitative descriptive inquiry of persistence* (Order No. 10809354). Available from ProQuest Dissertations & Theses Global. (2036952458). Retrieved from https://search.proquest.com/docview/2036952458

Elmer-Dewitt, P. (2016). Apple vs. FBI: What the polls are saying. *Fortune*. Retrieved from http://fortune.com/2016/02/23/apple-fbi-poll-pew

Endeley, R. E. (2018). End-to-end encryption in messaging services and national security - case of WhatsApp messenger. *Journal of Information Security, 9*(1), 95-99. https://doi.org/10.4236/jis.2018.91008

Fink, A. (2018). *How to conduct surveys*. A step-by-step guide. Thousand Oaks, CA: Sage Publications, Inc.

Herzberg, A., & Leibowitz, H. (2016). Can Johnny finally encrypt? Evaluating E2E-encryption in popular im applications. doi:10.1145/3046055.3046059.

Hilal, A. H., & Alabri, S. S. (2013). Using NVIVO for data analysis in qualitative research. *International Interdisciplinary Journal of Education, 2*(2), 181–186.

Jisha, K., & Jebakumar. (2014). A trend setter in mobile communication among Chennai youth. *IOSR Journal of Humanities and Social Science (IOSR-JHSS), 19*(9), 01-06. doi: 10.9790/0837-19970106

Jones, J., Turunen, H., & Snelgrove, S. (2016). Theme development in qualitative content analysis and thematic analysis. *Journal of Nursing Education and Practice, 6*(5), 100.

Kern, D. (2012). Understanding and implementing encryption backdoors. Retrieved from http://cse.ucdenver.edu/~dkern/CSC7002/paper.pdf

Levy, S. (2018, April). Cracking the crypto war. *WIRED*. Retrieved from https://www.wired.com/story/crypto-war-clear-encryption/ Lewis, J., Zheng, D., & Carter, W. (2017, February). The effect of encryption on lawful access to communications and data. *A report of the CSIS Technology Policy Program.*Washington, DC: Center for strategic and international studies

Max, Steven Patterson. (2016, April). WhatsApp copies apple's strong encryption defense. *Network World*, Southborough. Retrieved from https://search.proquest.com/docview/1779534877?accountid =44888

McCarthy, H. J. (2016). Decoding the encryption debate: Why legislating to restrict strong encryption will not resolve the "going dark" problem. *Journal of Internet Law. 20*(3). Retrieved from https://www.slideshare.net/HughJMcCarthy/decoding-the-encryption-debate-hugh-j-mccarthy-september-2016222905691pdf

Michalas, A. (2017). How WhatsApp encryption works - and why there shouldn't be a backdoor. The Conversation. Retrieved from https://theconversation.com/how-whatsapp-encryption- works-and-why-there-shouldnt-be-a-backdoor-75266

Novak, M. (2018). Paul Manafort learns that encrypting messages doesn't matter if the feds have a warrant to search your iCloud account. Retrieved from https://gizmodo.com/paul-manafort-learns-that-encrypting- messages-doesnt-ma-1826561511

Rashidi, Y., Vaniea, K., & Camp, J. (2016). Understanding Saudis' privacy concerns when using WhatsApp. doi:10.14722/usec.2016.23022.

Rastogi, N., & Hendler, J. (2017, June). WhatsApp security and role of metadata in preserving privacy. *Paper presented at the European Conference on Cyber Warfare and Security 269-XVI.* Dublin, Ireland.

Sagers, G., Hosack, B., & Rowley, R. (2015). *Where's the security in WiFi? An argument for industry awareness.* 48th Hawaii International Conference on System Sciences. IEEE Computer Society. Washington, DC, USA

Salkind, N. J. (2012). *Exploring research.* Upper Saddle River, NJ: Pearson Education, Inc.

Schneier, B., Seidel, K., & Vijayakumar, S. (2016). A worldwide survey of encryption products. *Berkman Center Research Publication 2016-2*. http://dx.doi.org/10.2139/ssrn.2731160

Shah, R. (2016). Law enforcement and data privacy: A forward-looking approach. *The Yale Law Journal, 125*(2), 326-559. Retrieved from http://digitalcommons.law.yale.edu/ylj/vol125/iss2/5

Sharma, Gaganpreet. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research 2017, 3*(7), 749-752.

SurveyMonkey Inc. (2019). Retrieved from www.surveymonkey.com. San Mateo, California, USA

Sutikno, T., Handayani, L., Stiawan, D., Riyadi, M. A., & Subroto, I. M. I. (2016). WhatsApp, Viber and Telegram which is best for instant messaging? *International Journal of Electrical and Computer Engineering, 6*(3), 909-914. doi: 10.11591/ijece.v6i3.10271

U.S. Census Bureau. (2015, June). Millennials Outnumber Baby Boomers and Are Far More Diverse, Census Bureau Reports. Release number CB15-113. Retrieved from https://www.census.gov/newsroom/press-releases/2015/cb15-113.html

U.S. Census Bureau. (2016, August). Number of IT workers has increased tenfold since 1970, census bureau reports. Retrieved from https://www.census.gov/newsroom/press- releases/2016/cb16-139.html

U.S. Department of Homeland Security. (2015, July). Going Dark: Encryption, Technology and the Balance between Public Safety and Privacy. *Senate Committee on the Judiciary.* Homeland Security Digital Library

U.S. Department of Labor. (2017). Bureau of Labor Statistics. Occupational employment statistics. Retrieved from https://www.bls.gov/oes/current/occ_state_lq_chart/occ_state_lq_chart.htm#

U.S. Department of Labor. (2019). Bureau of Labor Statistics. Labor force statistics from the current population survey. Retrieved from https://www.bls.gov/cps/cpsaat11b.htm

Vaziripour, E., Wu, J., Farahbakhsh, R., Seamons, K., O'Neill, M., & Zappala, D. (2018). *A survey of the privacy preferences and practices of iranian users of telegram*. Workshop on Usable Security (USEC). San Diego, CA. https://dx.doi.org/10.14722/usec.2018.23033

Wei, B., Doowon, K., Moses N., Yichen Q., Patrick G., & Michelle L. (2016). *An inconvenient trust: User attitudes toward security and usability tradeoffs for key-directory encryption systems*. Twelfth Symposium on Usable Privacy and Security. Denver, CO

WhatsApp (2017, December). WhatsApp encryption overview. *Technical white paper.* Retrievedfrom https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf

Yeboah, J., & Ewur, G. (2014). the impact of WhatsApp messenger usage on students performance in tertiary institutions in Ghana. *Journal of Education and Practice, 5*(6), 157-164. Retrieved from https://www.iiste.org/Journals/index.php/JEP/article/view/11241

