Quantum Malware

Lian-Ao Wu[(1)] and Daniel A. Lidar[(1,2)]
[(1)] *Chemical Physics Theory Group, Department of Chemistry, and Center for Quantum Information and Quantum Control, University of Toronto, 80 St. George St., Toronto, Ontario, M5S 3H6, Canada*
[(2)] *Departments of Chemistry, Electrical Engineering-Systems, and Physics, University of Southern California, Los Angeles, CA 90089*

When quantum communication networks proliferate they will likely be subject to a new type of attack: by hackers, virus makers, and other malicious intruders. Here we introduce the concept of "quantum malware" to describe such human-made intrusions. We offer a simple solution for storage of quantum information in a manner which protects quantum networks from quantum malware. This solution involves swapping the quantum information at random times between the network and isolated, distributed ancillas. It applies to arbitrary attack types, provided the protective operations are themselves not compromised.

I. INTRODUCTION

Quantum information processing (QIP) offers unprecedented advantages compared to its classical counterpart[1]. Quantum communication is moving from laboratory prototypes into real-life applications. For example, quantum communication networks ("quantum internet"[2]) have already been completed, and even commercialized[3]. Efforts to protect quantum information flowing through such networks have so far focused on environmental (decoherence) and cryptographic (eavesdropping) "attacks". Quantum error correction has been developed to overcome these disturbances[4,5,6].

Malware (a portmanteau of "malicious software"), familiar from classical information networks, is any software developed for the purpose of doing harm to a computer system[7]. This includes self-replicating software such as viruses, worms, and wabbits; software that collects and sends information, such as Trojan horses and spyware; software that allows access to the computer system bypassing the normal authentication procedures, such as backdoors, and more. In view of their strategic importance, when quantum information networks become widespread, it is likely that deliberately designed malware will appear and attempt to disrupt the operation of these networks or their nodes. We call the quantum version of these types of attacks *quantum malware*.

Quantum malware is a new category of attacks on quantum information processors. While it shares the "intelligent design" aspect of eavesdropping in quantum cryptography, one cannot assume that its perpetrators will attempt to minimally disturb a QIP task. Instead, while quantum malware will try to remain hidden until its scheduled launch, its attack can be strong and deliberately destructive. Moreover, generally it should be assumed that malware is able to attack at any point in time and target any component and part of the quantum devices in a quantum network. Quantum malware may appear in the form of a quantum logic gate, or even as a whole quantum algorithm designed and controlled by the attackers. In comparison with classical information processing, there are more ways to attack in QIP, because quantum states contain more degrees of freedom than their classical counterparts.

Here we propose a simple scheme to protect quantum memory in quantum information processors against a wide class of such malware. This scheme, while not foolproof, dramatically reduces the probability of success of an attack, under reasonable assumptions, which involve strengthening the defenders relative to the attackers. We note that if attacker and defender have exactly the same capabilities (including knowledge, e.g., of secret keys), a defense is likely to be impossible. Therefore, the question becomes, how much must one add to the defenders' capabilities, or subtract from the attackers', in order to have a secure network? The protocol we propose here to defend against quantum malware provides a possible answer to this question.

II. CAN QUANTUM MALWARE EXIST?

An early no-go theorem showed that it is not possible to build a fixed, general purpose quantum computer which can be programmed to perform an arbitrary quantum computation[8] . However, it is possible to encode quantum dynamics in the state of a quantum system, in such a way that the system can be used to stochastically perform, at a later time, the stored transformation on some other quantum system. Moreover, this can be done in a manner such that the probability of failure decreases exponentially with the number of qubits that store the transformation[9]. Such stochastic quantum programs can further be used to perform quantum measurements[10,11,12]. Thus it is entirely conceivable that quantum malware can be sent across a quantum information network, stored in the state of one or

more of the network nodes, and then (stochastically) execute a quantum program or measurement. Either one of these eventualities can be catastrophic for the network or its nodes. In the case of a maliciously executed measurement the outcome can be an erasure of all data. In the case of a quantum program one can imagine any number of undesirable outcomes, ranging from a hijacking of the network, to a quantum virus or worm, which replicates itself (probabilistically, due to the no-cloning theorem[13,14]) over the network.

III. QUANTUM MALWARE MODEL AND ASSUMPTIONS

While there is no limit to the number and character of possible malware attacks, they must all share the same fundamental characteristic: they comprise a set of elementary operations, "quantum machine-language", such as quantum logic gates and measurements. It is this simple observation, which also guided the early concept of the circuit model of quantum computing[15], that allows us to consider a general model of quantum malware, without resorting to specific modes of attack. We thus model quantum malware at this machine-language level. Clearly, this captures all "high-level" types of attack, since these must, by necessity, comprise such elementary operations. The operations can be unitary gates $U(t)$ driven by a time-dependent Hamiltonian $H(t)$, and/or measurements, taking place while a QIP task is in progress. We denote the series of malicious operations by the superoperator $\tilde{M}(\{|i\rangle\langle j|\}^{\otimes K})$, where $|i\rangle$ is an arbitrary basis state in the Hilbert space in which a qubit (one of K) is embedded. This notation includes measurements, as well as "leakage" operations that couple the two states of any qubit with the rest of its Hilbert space. For example, $\tilde{M}(\{|i\rangle\langle j|\})$ may have the structure of a quantum completely positive map[16]. This captures the most general type of quantum malware possible. The details of such quantum malware operations, i.e., the structure of $\tilde{M}(\{|i\rangle\langle j|\})$, are in general known only to the attackers, and we will not presume or need any such knowledge.

In order to protect against malware, in classical information processing one must assume that there is a means by which to determine, or at least estimate, a time interval δ within which the malware is off, so that malware-free data can be copied (backed up). For example, when one installs a firewall, or when one applies an anti-virus program, one must assume that these tasks themselves are malware-free. Similarly, we will assume that the quantum malware attack occurs in relatively short bursts, and that there are periods during which there is no attack. We note that it is in the interest of the attackers to remain hidden, or at least not to launch a continuous attack. For, otherwise, the defenders may simply decide that it is too risky to engage in any kind of activity, thus defeating the purpose of the attackers.

IV. NETWORK OPERATIONS PROTOCOL

Although the classical backup method is not directly applicable in the case of quantum information – because of the no-cloning theorem[13,14] – the basic idea of assuming malware-off periods while copying suggests an analogous mechanism for protecting quantum information against quantum malware. We note that the assumption that the attack is switched off every once in a while is not only reasonable for the sake of the adversary's purpose of maintaining an element of surprise, but is common also to quantum cryptography. For example, a probabilistic protocol for quantum message authentication (essentially a "secure quantum virtual private network") assumes that the sender and receiver are not subject to attacks by a third party at least while sending and measuring quantum states[17].

The protocol we describe below is deterministic and is designed to protect quantum information over time. The networks we consider comprise K nodes, which can either be the whole network or a part thereof. Each node contains a quantum computer. The network is used for the transmission of quantum information. Hence the nodes are connected via quantum and classical channels. The quantum channels are used for tasks requiring the transmission of quantum states, such as quantum cryptography[18]. The classical channels are useful, among other things for teleportation[19]. Henceforth, the terms "online"/"offline" applied to a network node mean that this node is connected/unconnected to the network. The defenders have access to three types of qubits, or quantum computers: (i) *Data qubits*, which can be either online or offline; (ii) *Decoy qubits*, which are online when the data qubits are offline, and vice versa; (iii) *Ancilla qubits*, which are always offline. In Table 1 we compare the assumptions we make about the respective capabilities of the defenders and attackers of the network. With the exception of the limitations listed in Table 1, the attackers are bound only by the laws of physics. Both defenders and attackers have access to clock synchronization,[20], which enables the defenders to make use of their set of secret network on-times.

One can envision any number of different methods by means of which the task of secure distribution of the network on-times to the defenders can be accomplished, including classical[21,22] and quantum secret sharing protocols[23,24,25], which are procedures for splitting a message into several parts so that no subset of parts is sufficient to read the message, but the entire set is. It is essential to the success of the protocol that only trusted parties are recipients.

The secret set $\{T_i\}$ is stored off-line by the defenders, and is never copied onto a computer that is accessed by the network. This provision is meant to preclude the attackers from ever gaining access to the times $\{T_i\}$, even if at some point they successfully (remotely) hijack a network node.

Capabilities of the defenders	Capabilities of the attackers
Have access to the secret set of network switch-on times $\{T_i\}_{i=1}^{M}$. This secret set is stored off-line by the defenders, and is never copied onto a computer that is accessed by the network.	Do not have access to the network switch-on times $\{T_i\}_{i=1}^{M}$, even if at some point they successfully (remotely) hijack a network node.
Can implement very fast communication across the network during the real (as opposed to decoy) network on-times.	Cannot discriminate between legitimate and decoy activity on the network.
Can quickly replacing the "data" quantum computers on their respective network nodes with "decoy" quantum computers.	Cannot interfere with the replacement of the data and decoy computer.

Table 1. Relative capabilities of defenders and attackers.

Our network protection protocol is given in Figure 1. After the preparatory steps (1) and (2), the protocol cycles through steps (3)-(6), with the next network on-times $\{T_i\}$ chosen from the previously distributed secret set. The protocol is further illustrated in Figure 2.

V. CONDITIONS FOR SUCCESS OF THE DEFENSE PROTOCOL

We note that if a malware attack ever takes place during the network-on times, replacement of data qubits by decoy qubits, decoy-reset, or SWAP operations, the protocol fails and the network must be completely reset. We can estimate the probability, p, of this catastrophic occurrence as follows. A reasonable strategy is to pick the times $\{T_i\}$ from a uniformly random distribution. The malware designers, on the other hand, may choose their attack interval times $\{\theta_j\}$ from some other distribution, not known to us. Let us characterize this latter distribution by a mean attack interval θ and mean attack length δ. Let the total time over which the protocol above is implemented be T. Let us also designate the operating times within a single cycle of our protocol by τ (i.e., $\tau = \tau_O + 2\tau_S + \tau_R$). Consider a particular attack window of length δ at some random time. The probability q_1 that the network is *off* during this window is $q_1 = [T - (\delta + \tau)]/T$, since there are two network-on intervals, one before and one after the attack window, and each must be a distance $\tau/2$ away from this window. In other words, the excluded interval is $\delta + 2(\tau/2)$. Now, since the network-on times are randomly distributed, the probability that this same attack window does not overlap any network-on interval, after M such intervals, is $q_M \approx q_1^M$ (this is an approximation since one should actually exclude overlapping intervals[26], but if the intervals are sufficiently sparse such overlaps can be neglected). The probability of at least one (catastrophic) overlap of this attack window with a network-on interval is $p = 1 - q_M$. Letting $M = cT$, where $0 < c < 1$ is a constant, we have $p \overset{T\to\infty}{\longrightarrow} 1 - \exp[-c(\delta + \tau)]$, so that as long as c and τ (under our control), and δ (under the attackers' control) are sufficiently small, we have $p \approx c(\delta + \tau) \gtrsim 0$. Another way of analyzing the optimal strategy is to note that there are, on average, a total of $A = T/(\theta + \delta)$ attack intervals, so that the expected number of catastrophic overlaps is $Ap = [T/(\theta + \delta)]\{1 - [1 - (\delta + \tau)/T]^M\}$, and this number must be $\ll 1$ for our protocol to succeed. Given an estimate of the attackers' parameters, θ and δ, and given that the state of technology will impose a minimum τ, we can use this result to optimize T and M. A simple estimate can be derived in the physically plausible limit $\delta, \tau \ll T$, where we can linearize the above expression and obtain the condition

$$M \ll (\delta + \theta)/(\delta + \tau). \tag{1}$$

(1) $t \le 0$: **Preparation step.** Everything offline. The state $\|\Omega(0)\rangle = \otimes_{i=1}^{K} \|\psi_i\rangle^d \|0_i\rangle^a \|0_i\rangle^\Delta$ is prepared locally. Superscripts d, a, and Δ denote data, ancilla, and decoy, respectively; i enumerates nodes, each of the states $\|\psi_i\rangle^d$ denotes an entire node data-state (of n_i data qubits), and each of the states $\|0_i\rangle^{a,\Delta}$ denotes an entire node ancilla or decoy-state (of n_i qubits). Local preparation means that all operations are done at the nodes, without any communication (classical or quantum) between nodes. Therefore at this point there is no risk of any malware infection and we can be confident that $\|\Omega(0)\rangle$ is a clean (malware-free) state.	(2) $0 \le t \le \tau_O$: **Real communication step.** The network is turned on and all data qubits are online. There is communication across the network, operations are performed, and qubits at different nodes may become entangled. The network evolves into the state $\|\Omega(\tau_O)\rangle = \|\Psi_1\rangle^d \|\mathbf{0}\rangle^a \|\mathbf{0}\rangle^\Delta$ Here $\|\mathbf{0}\rangle^{a,\Delta} = \otimes_{i=1}^{K} \|0_i\rangle^{a,\Delta}$ and $\|\Psi_1\rangle^d$ is the (possibly entangled) state of all data qubits in the entire network. We assume that τ_O is sufficiently short so that the malware has no time to interfere with the operation $\otimes_{i=1}^{K} \|\psi_i\rangle^d \mapsto \|\Psi_1\rangle^d$. I.e., we assume (since network-on times are secret and τ_O is short) that the state $\|\Psi_1\rangle^d$ is completely malware-free, and contains only legitimate information.	(3) $\tau_O \le t \le \tau_O + \tau_S$: **Short decoy step.** The network remains online, but real activity on the network is turned off and is replaced by decoy activity. Thus all data qubits from step (2) are taken offline, and are replaced by online decoy qubits, whose sole purpose is to fool the attackers into believing that the network is still active. A local SWAP operation, $\otimes_{i=1}^{K} S_i$, where $S_i \|x_i\rangle^d \|y_i\rangle^a = \|y_i\rangle^d \|x_i\rangle^a$, is performed between the data and ancilla qubits. This results in: $\|\Omega(\tau_O + \tau_S)\rangle = \|\mathbf{0}\rangle^d \|\Psi_1\rangle^a \left(\widetilde{M}_1 \|\mathbf{0}\rangle^\Delta\right)$, where $\|\mathbf{0}\rangle^d = \otimes_{i=1}^{K} \|0_i\rangle^d$. Note that we have allowed for malware to act on the decoy state via the operator $\widetilde{M}_1$. But, since malware has so far not had a chance to interact with the legitimate data, there is no risk of contamination of the ancilla qubits.
(4) $\tau_O + \tau_S \le t \le T_1 - \tau_S - \tau_R$: **Long decoy, "confuse the enemy" step.** The network remains online using the decoy qubits. During this long interval malware can continue to be transmitted across the network, store itself on the decoy qubits, or even attack the decoy qubits, via the operator $\widetilde{M}_2$: $\|\Omega(T_1 - \tau_S - \tau_R)\rangle = \|\mathbf{0}\rangle^d \|\Psi_1\rangle^a \left(\widetilde{M}_2 \widetilde{M}_1 \|\mathbf{0}\rangle^\Delta\right)$. Since the data has been swapped onto the ancillas, to which the malware has no access, the attack is harmless: the network state is safely stored in the ancillas (which may at this point be entangled across the network). Since data and ancilla qubits were entangled only during the execution of the SWAP gate, and we have assumed that the malware had no time to attack the data qubits prior to the SWAP operation, there is no mechanism for the malware to infect the ancillas. Conversely, our protocol fails if this assumption does not hold. To err on the side of caution, we may erase and hence reset the data qubits during this step.	(5) $T_1 - \tau_S - \tau_R \le t \le T_1$: **SWAP and reset step.** We perform another SWAP between the ancillas and data qubits (τ_S), then replace the decoy qubits by the data qubits so the latter are online, and reset the decoy qubits (τ_R). This erases any and all malware. We assume that the malware is not operating during this time. Then: $\|\Omega(T_1)\rangle = \|\Psi_1\rangle^d \|\mathbf{0}\rangle^a \|\mathbf{0}\rangle^\Delta$.	(6) $T_1 \le t \le T_1 + \tau_O$: **Real communication step.** The network is turned on and all data qubits are online again. We perform a new operation across the network: $\|\Omega(T_1 + \tau_O)\rangle = \|\Psi_2\rangle^d \|\mathbf{0}\rangle^a \|\mathbf{0}\rangle^\Delta$. We again assume that this operation is fast and that the time T_1 is unknown to the attackers, so that there is no malware attack.

FIG. 1: Network operations protocol.

If we further assume $\tau < \delta \ll \theta$, we find the intuitively simple result that *the number of network-on times cannot exceed the ratio of the mean attack interval to the mean attack length.*

We note that one might suspect that our protocol is in fact more vulnerable than suggested by the arguments above, given that an adversary might hijack quantum repeaters installed between network nodes and tweak the data (this scenario assumes quantum optical communication). However, we point out that there exists an alternative scheme to the use of quantum repeaters: in order to overcome photon decoherence and loss one may use a spatial analog of the quantum Zeno effect and "bang-bang decoupling", which involves only linear optical elements installed at regular

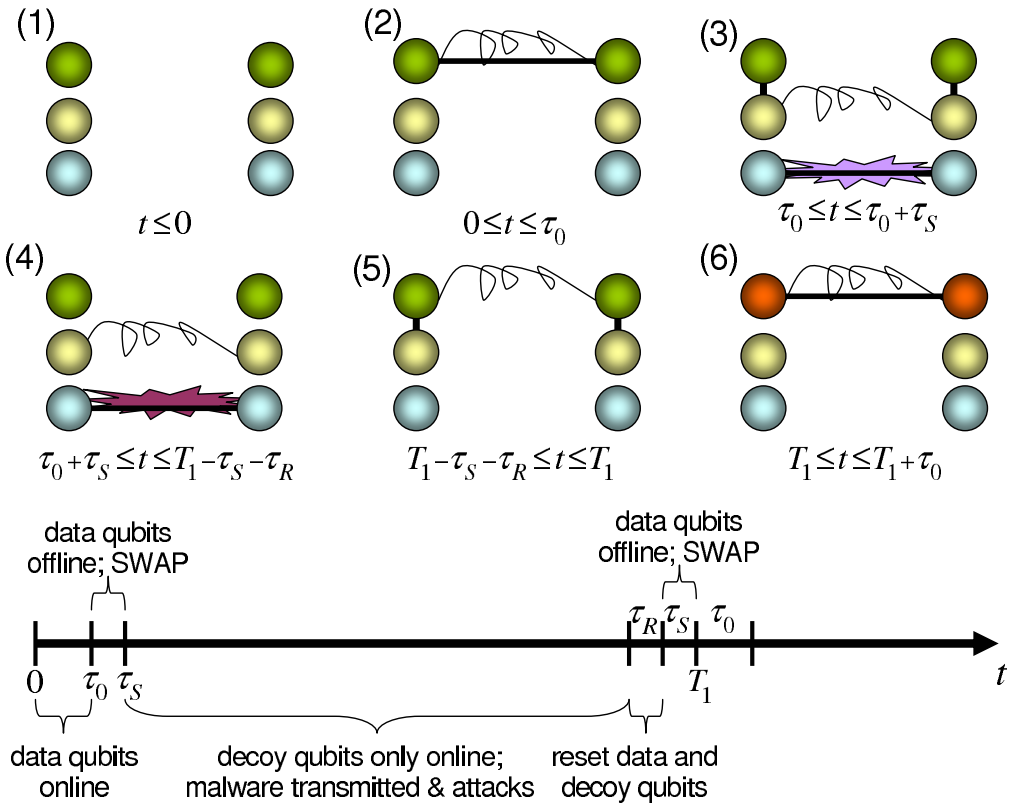

FIG. 2: Schematic of network operations protocol. Depicted in the top six parts is a simple network with $K = 2$ nodes and one data qubit in each node. For simplicity we do not depict other parts of quantum computers where the malware would reside. Parts (1)-(6) denote the first six steps in the protocol, starting from the first network cycle. The green dots (top) are data qubits, the yellow dots (middle) are ancillas, and the blue dots (bottom) are decoy qubits. Initially (1), the system is offline. When data qubits are connected by straight lines (2), the system is online. The curly lines (2) represent entangled qubits. The time at which the network is turned on is random and unknown to the malware makers, and the duration too short for them to interfere. In the ultrashort step (3) the network is off and the state of data and ancilla qubits is swapped, as represented by the vertical straight lines. The decoy qubits may be under attack. (4) Decoy qubits are subject to a malware attack. Whatever the attack, in (5) the data and decoy qubits are reset and the data qubits swapped with the ancilla qubits. Red data qubits (6) indicate the end of a network cycle, and the start of a new cycle. Bottom part: Timeline of the protocol.

intervals along an optical fiber[27]. Such a system cannot be hijacked because of its distributed nature. An attacker could at most remove, or tamper with, some of the linear optical elements, thus degrading the performance of the quantum noise suppression scheme.

VI. DETECTION OF AN ATTACK, AND INCREASING THE ROBUSTNESS OF THE DEFENSE PROTOCOL

A considerable improvement in the robustness of the stored quantum information is possible by replacing the SWAP operation with an encoding of each data qubit into a quantum error-detecting code[1]. Not only does this enable the application of quantum fault tolerance methods[4], *it also allows the defenders to check whether the data has been modified*, via the use of quantum error detection. However, since this does not allow us to change our assumptions about the relative weakness/strength of attackers and defenders, we do not here consider this possibility in detail.

We further note that it is possible to slightly relax the assumption that the malware makers cannot interfere during the real communication step (2). Indeed, it is possible to let the malware attack and/or store itself on another set of qubits connected to the network, as long as these qubits are not involved in storing the legitimate state being processed across the network. When executing the short decoy step (3), we must then assume that this other set of qubits does not interact with the ancillas.

VII. IMPLEMENTATION OF SWAP GATES

We now show how the SWAP gates needed in our protocol can be implemented in a variety of physical systems. Recall that above we distinguished between malware operating on the qubits' Hilbert space, and malware that includes operations on a larger Hilbert space ("leakage"). The implementation of S_i for malware $\tilde{M}(\{\vec{\sigma}_{i_d}\})$, where $\vec{\sigma}_{i_d}$ are the Pauli matrices on the data qubits i_d, without leakage, is direct. Assume that the Heisenberg interaction $\vec{\sigma}_{i_d}\cdot\vec{\sigma}_{i_a}$ between the ith data qubit and its ancilla is experimentally controllable, as it is in a variety of solid-state quantum computing proposals such as quantum dots[28]. Then the SWAP gate is $S_i = \exp(i\frac{\pi}{2}P_{i_d i_a})$, where $P_{i_d i_a} = \frac{1}{2}(\vec{\sigma}_{i_d}\cdot\vec{\sigma}_{i_a}+1) = \sum_{\alpha,\beta=0}^{1}(|\alpha\rangle_{i_d}\langle\beta|)\otimes(|\beta\rangle_{i_a}\langle\alpha|)$ is an operator exchanging between the ith data qubit and its ancilla, and the gate time τ_α ($\alpha = O, S, R$) is on the order of a few picoseconds[29]. The SWAP gate can be implemented in a variety of other systems, with other Hamiltonians, in particular Hamiltonians of lower symmetry[30,31].

If the attackers design malware capable of causing leakage into or from the larger Hilbert space with dimension N, the situation will be different. Generally, the malware superoperator $\tilde{M}$ is a function of transition operators of the form $|\alpha\rangle_i\langle\beta|$, where the case of $\alpha,\beta>1$ represents states other than the two qubit states $|0\rangle$ and $|1\rangle$. If both α and β are 0 or 1, the operation can be expressed in terms of Pauli matrices, e.g., $|0\rangle\langle 1| = \sigma^x + i\sigma^y$; if only one of either α or β is 0 or 1, the operation represents leakage to or from the qubit subspace. Let us define a generalized data-ancilla exchange operator, $P_{i_d i_a} = \sum_{\alpha,\beta=0}^{N-1}(|\alpha\rangle_{i_d}\langle\beta|)\otimes(|\beta\rangle_{i_a}\langle\alpha|)$. Then $P_{i_d i_a}|\alpha\rangle_{i_d}|\beta\rangle_{i_a} = |\beta\rangle_{i_d}|\alpha\rangle_{i_a}$ and $P^2_{i_d i_a} = I$, the identity operator. Therefore the generalized SWAP operator is

$$S_i = \exp(i\frac{\pi}{2}P_{i_d i_a}) = iP_{i_d i_a}, \tag{2}$$

and it follows directly that $S^\dagger\tilde{M}(\{i_d\})S = \tilde{M}(\{i_a\})$, where $S = \prod_i S_i^\dagger$. The exchange operator $P_{i_d i_a}$ can be implemented as a controllable two-body Hamiltonian in multi-level systems.

For a fermionic system such as excitonic qubits in quantum dots[32,33] or electrons on the surface of liquid helium[34], a qubit is defined as $|0\rangle = f_0^\dagger|\text{vac}\rangle$, $|1\rangle = f_1^\dagger|\text{vac}\rangle$, where $f_0^\dagger, f_1^\dagger$ are fermionic creation operators and $|\text{vac}\rangle$ is the effective vacuum state (e.g., the Fermi level). The most general attack uses an operator (a Hamiltonian or measurement) that can be expressed in terms of $F_{i_d} \equiv (f_{0,i_d}^\dagger)^k(f_{1,i_d}^\dagger)^l(f_{0,i_d})^m(f_{1,i_d})^n$ (where k,l,m,n are integers) acting on the ith data qubit. These operators can be shifted to the corresponding ancilla, via control of a two-body fermionic Hamiltonian. Namely, $S_i^\dagger F_{i_d} S_i = F_{i_a}$, where the SWAP operator for the ith fermionic particle reads $S_i = S_i^0 S_i^1$, where

$$S_i^q = \exp[\frac{\pi}{2}(f_{q,i_d}^\dagger f_{q,i_a} - f_{q,i_a}^\dagger f_{q,i_d})],$$

with $q = 0, 1$. This can be proven easily using the identities

$$\begin{aligned} e^{-\phi(f_d^\dagger f_a - f_a^\dagger f_d)} f_d^\dagger e^{\phi(f_d^\dagger f_a - f_a^\dagger f_d)} &= \cos\phi f_d^\dagger + \sin\phi f_a^\dagger, \\ e^{-\phi(f_d^\dagger f_a - f_a^\dagger f_d)} f_d e^{\phi(f_d^\dagger f_a - f_a^\dagger f_d)} &= \cos\phi f_d + \sin\phi f_a, \end{aligned}$$

which follow from the Baker-Hausdorff formula $e^{-\alpha A}Be^{\alpha A} = B - \alpha[A,B] + \frac{\alpha^2}{2!}[A,[A,B]] - ....$ The relation $S_i^\dagger F_{i_d} S_i = F_{i_a}$ implies that the action of any "fermionic malware" is shifted by the SWAP gate from the data to the ancilla particle. The very same construction works also for bosonic systems, such as the linear-optics quantum computing proposal[35]. There a qubit is defined as $|0\rangle = b_0^\dagger|\text{vac}\rangle$, $|1\rangle = b_1^\dagger|\text{vac}\rangle$, where $b_0^\dagger, b_1^\dagger$ are bosonic creation operators. The relations we have just presented for fermions hold also for bosons, provided one everywhere substitutes bosonic operators in place of the fermionic ones.

VIII. CONCLUSIONS

What sets quantum malware apart from the environmental and eavesdropping attacks is that the latter are typically weak (in the sense of coupling to the quantum information processing (QIP) device), while the former can be arbitrarily

strong, can attack at anytime, and can target any part of a quantum device. Indeed, a malicious intruder, intent on disrupting information flow or storage on a quantum network, will resort to whatever means available. In contrast, the QIP-environment interaction will be a priori reduced to a minimal level, and an eavesdropper will attempt to go unnoticed by the communicating parties. For this reason one cannot expect quantum error correction to be of use against quantum malware, as it is designed to deal with small errors. The same holds true for quantum dynamical decoupling[36] or other types of Zeno-effect like interventions[37]. Decoherence-free subspaces and subsystems[38], on the other hand, do not assume small coupling, but do assume a symmetric interaction, which is unlikely to be a good assumption in the case of quantum malware. We further note that of all possible types of quantum malware, as far as we know only quantum trojan horses have been considered previously, in the quantum cryptography literature. In particular, in the context of the security proof of quantum key distribution, it was shown that teleportation can be used to reduce a quantum trojan horse attack to a classical one[39]. Finally, we note that the attacks we are concerned with are on the quantum data, not the quantum computer software; the latter is generally itself a list of classical instructions, and can be cloned.

Experience with classical information processing leaves no doubt that the arrival of quantum malware – malware designed to disrupt or destroy the operation of quantum communication networks and their nodes (quantum computers) – is a matter of time. When this happens, overcoming the problem of quantum malware may become as important as that of overcoming environment-induced decoherence errors. In this work we have raised this specter, and have offered a relatively simple solution. Our solution invokes a network communication protocol, wherein trusted parties operate the network at pre-specified times, and quickly swap the information out of the network onto a quantum backup system. Such a protocol slows the network down by a constant factor, and therefore does not interfere with any quantum computational speedup that depends on scaling with input size. The success of our protocol depends strongly on the ability to perform very rapid swapping between data and ancilla qubits. This suggests the importance of the design of fast and reliable swapping devices. This can be done for a variety of physical systems, as shown above. As long as the swapping can be done sufficiently fast, and as long as there exists a mechanism for secure distribution of the network on-times only among trusted parties, we have shown that the quantum network will be unharmed by a very general model of quantum malware. On the other hand, if these assumptions are not satisfied and an attack is successful, one must unfortunately reset the network, pending the development of a "quantum anti-virus program" that would clean infected data. The latter is a very interesting open research problem. Our protocol is similar to "paranoid" classical protocols employed in military systems that are under attack, which are shut down a great deal of the time, and then are suddenly opened up in order to perform a useful task. However, there is a distinct quantum aspect to our protocol, which is that it preserves entanglement across the network. In this sense our protocol, while being a conceptually simple generalization of established classical methods, offers a genuine step forward towards quantum network security against quantum malware.

Acknowledgments

Financial support from the DARPA-QuIST program (managed by AFOSR under agreement No. F49620-01-1-0468) and the Sloan Foundation (to D.A.L.) is gratefully acknowledged. We thank Dr. Y. Xu (Microsoft Research Asia, Beijing) and Prof. Hoi-Kwong Lo (University of Toronto) for helpful discussions.

[1] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
[2] J.P. Dowling and G.J. Milburn, Phil. Trans. Roy. Soc. (Lond.) **361**, 1655 (2003).
[3] C. Elliott, eprint quant-ph/0412029.
[4] A.M. Steane, Nature **399**, 124 (1999).
[5] R. Cleve, D. Gottesman, H.-K. Lo, Phys. Rev. Lett. **83**, 648 (1999).
[6] P.W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).
[7] See, e.g., http://en.wikipedia.org/wiki/Malware.
[8] M.A. Nielsen and I.L. Chuang, Phys. Rev. Lett. **79**, 321 (1997).
[9] G. Vidal, L. Masanes, and J. I. Cirac, Phys. Rev. Lett. **88**, 047905 (2002).
[10] M. Rosko, V. Buzek, P.R. Chouha, and M. Hillery, Phys. Rev. A **68**, 062302 (2003).
[11] J. Fiurasek and M. Dusek, Phys. Rev. A **69**, 032302 (2004).
[12] G.M. D'Ariano and P. Perinotti, Phys. Rev. Lett. **94**, 090401 (2005).
[13] W.K. Wootters and W.H. Zurek, Nature **299**, 802 (1982).
[14] D. Dieks, Phys. Lett. A **92**, 271 (1982).

[15] D. Deutsch, Proc. Roy. Soc. London Ser. A **425**, 73 (1989).
[16] K. Kraus, *States, Effects and Operations*, *Fundamental Notions of Quantum Theory* (Academic, Berlin, 1983).
[17] H. Barnum *et al.*, in *Proc. 43rd Annual IEEE Symposium on the Foundations of Computer Science (FOCS '02)* (IEEE Press, 2002).
[18] D. Gottesman, H.-K. Lo, Physics Today **53**, 22 (2000).
[19] C.H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres and W.K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
[20] V. Giovannetti, S. Lloyd, S., L. Maccone, Nature **412**, 417 (2001).
[21] G. Blakely, Proc. of the AFIPS National Computer Conference **48**, 313 (1979).
[22] A. Shamir, Comm. Assoc. Comput. Mach. **22**, 612 (1979).
[23] D. Gottesman, Phys. Rev. A **61**, 042311 (2000).
[24] M. Hillery, V. Buzek, and A. Berthiaume, Phys. Rev. A **59**, 1829 (1999).
[25] A. Karlsson, M. Koashi, N. Imoto, Phys. Rev. A **59**, 1999 (162).
[26] D.A. Hamburger, O. Biham and D. Avnir, Phys. Rev. E **53**, 3342 (1996).
[27] L.-A. Wu and D.A. Lidar, Phys. Rev. A **70**, 062310 (2004).
[28] D. Loss and D.P. DiVincenzo, Phys. Rev. A **57**, 120 (1998).
[29] G. Burkard, D. Loss and D.P. DiVincenzo, Phys. Rev. B **59**, 2070 (1999).
[30] D.A. Lidar and L.-A. Wu, Phys. Rev. Lett. **88**, 017905 (2002).
[31] N.E. Bonesteel, D. Stepanenko, and D.P. DiVincenzo, Phys. Rev. Lett. **87**, 207901 (2001).
[32] P. Chen, C. Piermarocchi, and L. J. Sham, Phys. Rev. Lett. **87**, 067401 (2001).
[33] E. Biolatti, R.C. Iotti, P. Zanardi, and F. Rossi, Phys. Rev. Lett. **85**, 5647 (2000).
[34] P.M. Platzman and M.I. Dykman, Science **284**, 1967 (1999).
[35] E. Knill, R. Laflamme, and G.J. Milburn, Nature **409**, 46 (2001).
[36] L. Viola, J. Mod. Optics **51**, 2357 (2004).
[37] P. Facchi, S. Tasaki, S. Pascazio, H. Nakazato, A. Tokuse, D.A. Lidar, Phys. Rev. A **71**, 022302 (2005).
[38] D.A. Lidar, K.B. Whaley, in *Irreversible Quantum Dynamics*, Vol. 622 of *Lecture Notes in Physics* (Springer, Berlin, 2003), p. 83. Eprint quant-ph/0301032.
[39] H.-K. Lo and H.F. Chau, Science **283**, 2050 (1999).