Quantum Computational Supremacy: Security and Vulnerability in a New Paradigm

Deborah Brennan

Technological University Dublin, deborah.brennan@tudublin.ie

Follow this and additional works at: https://arrow.tudublin.ie/icr

Part of the Communication Technology and New Media Commons

Recommended Citation

Available at: https://arrow.tudublin.ie/icr/vol16/iss1/10

This Article is brought to you for free and open access by the Journals Published Through Arrow at ARROW@TU Dublin. It has been accepted for inclusion in Irish Communication Review by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License
Quantum computational supremacy: 
Security and vulnerability in a new paradigm

Deborah Brennan

Abstract

Despite three decades of research, the field of quantum computation has yet to build a quantum computer that can perform a task beyond the capability of any classical computer – an event known as computational supremacy. Yet this multi-billion dollar research industry persists in its efforts to construct such a machine. Based on the counter-intuitive principles of quantum physics, these devices are fundamentally different from the computers we know. It is theorised that large-scale quantum computers will have the ability to perform some remarkably powerful computations, even if the extent of their capabilities remains disputed. One application, however, the factoring of large numbers into their constituent primes, has already been demonstrated using Shor’s quantum algorithm. This capability has far reaching implications for cybersecurity as it poses an unprecedented threat to the public key encryption that forms an important component of the security of all digital communications. This paper outlines the nature of the threat that quantum computation is believed to pose to digital communications and investigates how this emerging technology, coupled with the threat of Adversarial Artificial Intelligence, may result in large technology companies gaining unacceptable political leverage; and it proposes measures that might be implemented to mitigate this eventuality.
Introduction

Scholarship in communications and media literacy has advanced into the 21st century too often by merely appending the word ‘digital’ to its categories of interest, and by adopting software-based tools to conduct some of its research. While long-term continuities in the issues affecting media production and dissemination are undoubtedly important, and there has been no shortage of investigation based upon the peculiar modes and interactions of the online communications environment, there has been a tendency for research to stop outside the door of computation itself, so that even the all-important algorithms that govern the behaviour of social media platforms are more likely to be mentioned in passing than probed in depth. Such relative indifference to computation would be more justifiable if we could assume, as many writers do, that technological advances in computing occur at a steady and predictable rate as the commercial focus of computation shifts to data analytics and methods of artificial intelligence (AI) such as machine learning (ML) and natural language processing (NLP). This assumption, however, ignores significant advances in the field of quantum computation. As one important scholar in this area stated in March 2018: ‘Whoever can build a fully functioning quantum computer will rule the world’ (Soo, 2018). In this paper, I will interrogate this proposition by exploring the ramifications of these advances for online security and further corporate control of the Internet. The sections on quantum computers and P vs NP contain elements of quantum physics and mathematics, respectively, which I have attempted to present as clearly as possible with no prior knowledge required of the reader. I hope that the reader may enjoy these sections – however, the paper may be read without them.

Quantum computers

Despite three decades of research, as of 2018 quantum computation remains a multi-billion dollar industry that has yet to produce a working prototype with more than around 8 operational Bytes. Potential quantum technologies are fraught with issues deriving from the science on which they depend, nonetheless, all of the major technology companies and a large number of governments and research
institutes continue to invest heavily in the field. Google, IBM and Microsoft along with the Chinese technology companies Alibaba and Tencent all have quantum or hybrid quantum-artificial intelligence (AI) laboratories with large research budgets. Government investments have also been generous, with the European Commission, for example, pledging €1 billion for research into quantum technologies (European Commission, 2016). The stated goal of organisations engaging in research in quantum computation is often the achievement of computational supremacy, the point where a quantum computer performs a computational task that is beyond the capability of any classical computer – literally a paradigm-changing event for information technology.

Quantum computers are entirely different in concept and operation from the computers we commonly use. Quantum devices operate according to two key postulates of quantum physics: superposition and entanglement. Superposition means that each quantum bit, or qubit, can represent both 1 and 0 at the same time whereas the more familiar bit must take either the value of 1 or 0. Theoretically, a qubit in superposition can hold an infinite amount of information which can be manipulated using quantum gates finally to yield a value of either 0 or 1 on measurement\(^2\). It is theorised that this is one of the sources of the power of quantum computation. Entanglement means that qubits in a superposition can be correlated with each other allowing them to work together to facilitate something like massive parallel processing on a single device\(^3\). Quantum computers will exploit these properties to solve some problems that are considered very difficult or impossible using classical computers, and they will solve some of these problems at incredibly high speeds. The rewards for working hardware running novel algorithms from this new paradigm are expected to be very high, with promises of technologies offering considerable advances in fields such as artificial intelligence, molecular simulation, hyperreal gaming and stock market prediction.

\(^2\)This is according to the so-called Copenhagen Interpretation of quantum mechanics via the so-called *collapse of the wave function*; the *Multiverse or Many Worlds* interpretation offers a different explanation. See, for example, Vaidman (2014)

\(^3\)The mechanism by which this speedup occurs is still disputed and its nature has far reaching consequences for both quantum physics and quantum information theory. It has been suggested that the speedup lies either in quantum dynamics (Schrödinger equation) or in the quantum state itself (the wave function, \(\Psi\)).
The extraordinary properties of entangled quantum bits in superposition mean that even a very small number can perform remarkable computation. The most important quantum algorithm to date, the factoring algorithm due to Peter Shor (1994), is considered capable of factoring large integers (numbers) to their constituent primes by creating a what Pitowsky terms a ‘clever superposition’ (2002) of entangled qubits and extracting a solution in a short, or polynomial, time frame. On a functioning 100 bit quantum computer, Shor’s algorithm could break RSA, the most commonly used public key encryption protocol on the Internet, in hours to days. Scaling up to Quantum Kilo Bytes, RSA public key encryption becomes completely ineffective.

Shor’s algorithm belongs to a very significant class of quantum algorithms known as the Hidden Subgroup Problem (HSP). Variants of this problem have been discovered that can, in theory, solve the mathematics behind RSA, ECDSA and DSA (Grosshans et al., 2015), all of the main public key encryption protocols in current use on the Internet. Clearly, quantum technologies pose a very real threat to current information security.

**The nature of information security**

There are many ways to conceptualise information security and this paper does not intend to detail these to any great degree; rather it endeavors to convey the idea that security needs to be complete, that every component of information security is a potential weak point that can be exploited regardless of the strength of the other components, much like how an open window in an otherwise secure building constitutes a weakness in its security. For the purposes of this paper the simple, but widely used, principles of the CIA Triad of Information Security – Confidentiality, Integrity, and Availability (Lefkovitz et al., 2017) – will be sufficient. Ensuring a high degree of confidentiality requires, but is not limited to, employing measures to actively ensure that sensitive information is not intercepted by unauthorized parties while in transit or in storage, while ensuring that resources

---

4 Shor’s algorithm has been run successfully on a true quantum device and has demonstrated the factorisation of small numbers, but the principle is proven for large numbers although there may be issues with scaling.
5 The CIA Triad is a widely adopted information security benchmark model used to evaluate the information security of an organization, other models may be used to model individual aspects of security e.g. the PAIN (Privacy, Authority, Integrity & Non-Repudiation) model for cryptography.
are available to intended parties. Integrity means that data must not be altered in transit or in storage without authorization; among other measures this involves both symmetric (private key) and asymmetric (public key) encryption protocols. Ensuring availability includes security against denial of service (DoS) attacks that consume the entirety of a network’s resources, making them unavailable to legitimate users. Because of their threat to public key encryption, Shor’s and similar HSP algorithms are known to pose a risk to two of the Triad’s components, confidentiality and integrity. It is also thought that other quantum algorithms, such as variants of Grover’s algorithm (1996), may be used to exploit weaknesses in the implementation of some private key schemes\textsuperscript{6}.

Even without the use of quantum technologies, systems are under constant attack. Absolute security of data is impossible to achieve (and sometimes not desired\textsuperscript{7}) and the cost increases greatly with the degree of security provided. A large proportion of costs incurred is due to the higher processing overheads required for stronger encryption; the other large cost is security expertise\textsuperscript{8}. Different levels of security are provided for different internet and cloud services; for example, security policies for Internet banking are stricter than those for social media sites and messaging systems. A balance is always struck on the basis of the conflicting constraints of security and performance\textsuperscript{9}, legal constraints and the value of the data to be protected. Although security systems vary greatly in terms of composition and policy, practically all use the same or very similar encryption protocols and these protocols all face the same risks, real and potential, from quantum computers\textsuperscript{10}.

**The P vs. NP conundrum**

Mitigating the threat of quantum computers has been largely reduced to finding replacement algorithms for those behind the Public Key Protocols in current use, and this is not only because this is where the imminent threat is widely held to lie.

\textsuperscript{6} There may be other, as yet undiscovered, quantum algorithms which could threaten these same, and other, aspects of information security.

\textsuperscript{7} This point is supported by the classification of cryptographic technologies as munitions under US law until 1992; certain restrictions remain under international agreements.

\textsuperscript{8} Other costs include the cost of updating to current versions of software and software licences for firewalls and other security specific softwares.

\textsuperscript{9} Strong encryption slows performance and can make system unacceptably slow for end users.

\textsuperscript{10} Computer systems face very many other security threats; this will be discussed later.
Despite early successes in the field, quantum information theory has produced only a few algorithms, including Shor’s, that have been demonstrated capable of exponential speed up over their classical counterparts (Montanaro, 2015). Furthermore, many researchers are of the view that the once promised ubiquitous parallel processing power of quantum algorithms is in reality only possible in a very limited set of cases (Aaronson, 2013). The problem, they claim, comes down to the mathematics, specifically the computational complexity class of the mathematical problems under consideration.

Loosely speaking\(^{11}\), computational complexity is a measure of how much computational resource (time, energy, etc) it takes a computer to find a solution to a mathematical problem and in particular, the manner in which the resource requirement grows as input size grows (e.g. polynomially or exponentially). In computational complexity theory, class P (polynomial time) are relatively easy problems for computers so, on average, take few resources or short time to solve. For an average input, class NP (nondeterministic polynomial time) problems take a lot more computational resources to solve and so a lot more time\(^{12}\). Rarely, a problem can be thought to be in NP but turns out to also be in P when a solution is found that is ‘simpler’ than expected. It can also be unclear to which class(es) a particular problem belongs\(^{13}\).

In order to solve difficult problems in NP\(^{14}\) in a short time frame (or, equivalently, with few resources) these problems would need to reduced or broken down to a set of simpler problems in P. It is conjectured, but not proven, that this is not possible (Gill, 1977). The question is described in complexity theory as P vs NP and has great significance for computation in general\(^ {15}\). It is argued that the capabilities of quantum computers are limited by the conjecture that P does not equal NP (Aaronson, 2013), that the class of problems in NP that are not also in P

---

\(^{11}\) For understandability, what is presented is really a description of computational difficulty which is in many ways analogous to computational complexity.

\(^{12}\) Individual instances of a problem type can take significantly longer or shorter time to solve.

\(^{13}\) There are more classes in the computational complexity hierarchy. This discussion is limited to P and NP for clarity and understandability and as P vs NP may be significant in quantum information theory.

\(^{14}\) These problems are not also in the P complexity class, so do not have polynomial time solutions.

\(^{15}\) The P vs NP is highly important in complexity theory, for both classical and quantum computing, it was chosen as one of the 7 most significant unsolved mathematical problems by the Clay Mathematics Institute, The Millennium Problems (Clay Mathematics Institute, 2000).
cannot be reduced to a set of simpler problems that can be solved easily and quickly, even by a quantum computer\textsuperscript{16}. A subset of the NP class that is of particular interest to the field of quantum information theory, and any threat it poses to information security, is the NP-Complete class. NP-Complete problems\textsuperscript{17} are a set of problems for which finding a solution to any one problem will also provide a solution to all other problems in the NP-Complete set. In other words, if a polynomial time solution is found for any one NP-Complete problem, it is found for all NP-Complete problems.

The mathematical problem of factoring products of large prime numbers into their constituent parts, that problem that lies behind the security of the RSA security protocol, is likely to be in class NP\textsuperscript{18} but very significantly not in NP-Complete. Consequently, the fact that Shor’s algorithm has been successfully ‘demonstrated’ on quantum hardware has no predictive value for the success or otherwise of quantum algorithms in general. Still, regardless of how widely held the conjecture that $P \neq NP$ is, it remains conjecture. Furthermore, quantum algorithms differ significantly in structure and execution from their classical counterparts and, perhaps more importantly, are hardware dependent, which is not the case for most classical algorithms. For these reasons, the complexity of quantum algorithms is measured according to a different complexity class system which does not map directly to the classical system. For example, BQP and QMA are generally considered to be the bounded-error quantum analogues\textsuperscript{19} of the classical complexity classes $P$ and $NP$ respectively (Aaronson, 2009); however, Shor’s algorithm is believed to belong to NP in the classical system but not QMA in the quantum system, reflecting the fact that there exists no known classical algorithm that can factor large integers to their constituent prime factors in polynomial (or short) time.

There remains much to be reconciled and understood in the mathematics and physics that underpin the paradigm of quantum computing. The opinions of those considered experts in the field vary widely: for example, Scott Aaronson, a

\textsuperscript{16} The significance of NP-Complete complexity for quantum information theory will be discussed later.
\textsuperscript{17} Problem here means ‘problem type’ e.g. The Knapsack Problem or the Graph Colouring Problem not individual instances of that problem.
\textsuperscript{18} Prime Factoring is also believed to be in both $P$ and $CoNP$ but is not considered to be NP-Complete so that Shor’s algorithm does not solve the $P$ vs $NP$ problem.
\textsuperscript{19} For discussion see, for example, Younes & Rowe (2015).
prominent quantum computation scholar, holds that \( P \neq NP \) indeed sharply limits the possibilities of quantum computers; another prominent scholar, John Preskill, holds that quantum computers will themselves accelerate the development of quantum algorithms, addressing problems in ways that may not be immediately explicable; and David Deutch, a founder of the field, contends that the theory of quantum computation can be generalised to all physical processes in a complex multiverse, with classical computing explained as a special instance within a quantum paradigm.

It is not disputed, however, that Shor’s algorithm itself is mathematically possible and verified, meaning that it works in principle. It has been demonstrated in practice for small integers (numbers) on working prototypes of quantum computers and so it is reasonable to think that the factoring of large numbers into their constituent primes is a possibility that may continue to become easier, cheaper and more available as quantum technologies mature, making quantum computing a legitimate concern for information security in the near to medium term.

Currently, almost all security on the Internet uses the type of encryption that could be broken by a quantum computer running Shor’s algorithm as part of one or more protocols; this includes all app messaging encryption, email, current browser security, Internet banking and logins to cloud resources. Such evident risk might suggest we change our security protocols to ones that rely on NP-Complete problems, that is to mathematical problems to which no solution algorithm provides any significant speed up over trying every possible solution until the correct answer is found. These are problems that most mathematicians believe to be impossible to solve easily for all cases, in the classical paradigm at least.

There are a number of difficulties with this approach: public key cryptography, as it is currently conceived, relies on mathematical problems that are hard to solve every time no matter what the numbers are, otherwise some keys would be easy to discover or ‘crack’ and, in an operational security protocol, it would likely be impractical to ‘filter out’ these weak keys. Many, if not most, known problems in NP-Complete have so called ‘easy instances’ ruling them out as candidates (Talbot and Welsh, 2006). Another requirement for public key encryption systems is that the mathematical problem has an intentional ‘hidden trap door’ which effectively means that if a party has a key, they can easily decrypt a message. (This property
is often erroneously referred to as ‘being a one-way function’. The existence of true one-way functions has not been proven and if it were, it would prove that $P \neq NP.$ (Hartmanis & Hemachandra, 1999). In addition to these two primary requirements for an ideal post-quantum encryption protocol, there are many other requirements around resources, implementation and practical integration with existing systems and protocols. In short, finding suitable candidate algorithms is enormously challenging.

**The search for post-quantum algorithms**

In response to the potential threat that the quantum computing paradigm poses to information security, in December 2016 the US National Institute for Standards in Technology (NIST) issued a call for a first round of proposals for new so called quantum resistant algorithms to be used in the development of cryptographic standards (NIST, 2016). The standards are to be published as Federal Information Processing Standards (FIPS) or Special Publications (SPs). NIST invites and encourages participation from the cryptographic community as well as the general public in a process of finding a number of candidate replacement algorithms. The first of what is expected to be several rounds of submissions closed in November 2017 with 69 submissions accepted, of which five have subsequently been withdrawn.

NIST's approach is cautious, recognising that ‘the current scientific understanding of the power of quantum computers is far from comprehensive’ (NIST, 2017) and that candidate solution algorithms may be based on significantly different mathematics and design from those in current use. The organisation anticipates that the evaluation process may be ‘significantly more complex’ than the evaluation of the SHA-3 Hash algorithm candidates (Alshaikhli et al., 2012), for example, a process which took about eight years from call for proposals to official release of protocol details. This is without any acceptance period (where the community gains trust in the algorithm through its resistance to attack),

---

20 FIPS comprises 4 security levels and is the de facto international standard for information security prescribing not only cryptography but also security policy and hardware measures (e.g. tamper evident enclosures and true atomic-decay random number generation).
commercial implementation and roll-out, processes which together may take the same number of years again.\textsuperscript{21}

NIST has also stated that it believes that the transition should happen well in advance of the appearance of large scale quantum computers ‘so that any information that is later compromised by quantum cryptanalysis is no longer sensitive when that compromise occurs’ (NIST, 2016). This is interesting on two counts, the first being that there exist current trusted cryptographic algorithms for data storage that are believed to be resistant to all currently known quantum algorithms (NIST, 2018) (even those which have never been implemented in hardware), implying that NIST expects that the field of quantum computation to produce more and different algorithms. This would suggest that NIST’s scientists do not put complete faith in any notion that quantum computing is intrinsically limited by either technical problems or the $P \neq NP$ conjecture. The second, and perhaps more significant point is the loose terminology of ‘large scale’ as opposed to the usual (to the literature) ‘universal’\textsuperscript{22} quantum computer.

It has been generally considered that the goal of the field should be to produce some type of quantum analogue to what we currently consider a computer to be, a universal quantum computer, but this is unlikely to be the path that quantum computing takes. It would make little sense, for example, not to take advantage of the monumental advances in the field of AI in the development of new quantum technologies, and since algorithms from the classical and quantum paradigms are fundamentally different, in both concept and execution, a hybrid solution is the most likely possibility. This idea is not new: parts of Shor’s algorithm are classical in nature\textsuperscript{23} and the off-loading of compute-intensive operations to dedicated devices such as GPUs for graphics and ASICs for cryptographic routines, for example, is commonplace in current technologies. NIST itself states that it is aware that groups are developing hybrid cryptographic schemes although it is not considering such systems at present (NIST, 2017). In Europe, Europol echoes NIST’s views, citing European Union funding for research into post-quantum

\begin{enumerate}
\item See Bitcoin Forum (2013) for example, a forum discussion on the proposed use of SHA-3 in Bitcoin.
\item The mathematical model for a ‘universal’ computer also known as the Turing machine (Turing, 1936).
\item The first part of the Shor’s algorithm converts the factoring problem into the problem of finding the period of a function, this is implemented classically.
\end{enumerate}
algorithm candidates and warning that industry needs to keep up to date with developments if security is to be maintained (Europol, 2016).

It is not denied by NIST and other agencies concerned with information security that quantum technologies, alone or as part of hybrid quantum-classical systems, pose a real and largely unquantifiable threat in the medium to long term. The question remains, however, if the response to this threat is adequate and appropriate.

**Information security in practice**

The information security industry has developed and matured around an attack-fix cycle in which some, but significantly not all, security attacks are detected and analysed and appropriate 'fixes' are applied directly to systems or pushed out to end users as part of software upgrades. Software and its security then continue to operate as intended until the next attack and fix, or the next routine upgrade. Attacks, when detected, are usually dealt with promptly and are rarely publicised either because of their relative insignificance or because of fears of the damage to trust they can cause. The detection and mitigation of security attacks is complex and challenging but nonetheless generally relatively routine for most mature organisations, with resources allocated to security proportional to the sensitivity of information to be protected.

Nevertheless, no system is ideal and compromises are made to balance conflicting requirements such as security and system response time, and this can mean exposing the system to risk. For instance, 'light' encryption has been sometimes used to avoid the computational overheads of RSA and similar protocols and here the algorithms themselves may be the focus of attack. In June 2012, the business network LinkedIn was targeted in a cyberattack in which the passwords of more than 6 million users were stolen (Perlroth, 2018). It has been widely speculated, that the passwords had been protected using an 'unsalted' Hash algorithm\(^\text{24}\) such as SHA-1 or similar, a type of encryption known, even then (Theocharoulis et al.,

\(^{24}\) Unsalted hashes mean that when two or more users choose the same password, the same hash is generated each time. In this scheme, an attacker who knows the hash for a given password, can find the password whenever it is chosen by a new user. Hence commonly used or 'meaningful' (e.g. Italia90) passwords are easy to crack when unsalted hashes are used.
2010), to be vulnerable to so called rainbow attacks. Some data from the attack were posted to a Russian hacker site soon after and finally, in 2016, the full data were offered for sale on the so called dark web (Mathews, 2017). Despite the obvious vulnerability of SHA-1 and similar algorithms exposed by the 2012 LinkedIn attack, the Internet company Yahoo was the target of a similar attack on its unsalted MD5 Hash-protected user data the following year, when 3 billion user accounts were compromised (Stempel and Finkle, 2018). Interestingly, this attack went undetected for some time until it was eventually discovered during an investigation into a subsequent attack which took place in 2014, when 500 million user accounts were compromised. Yahoo claims that by the time of the second attack it had moved ‘the majority’ of its accounts to the protection of stronger encryption (Stempel and Finkle, 2018) offered by the more secure, but resource-intensive, BCrypt algorithm (Alabaichi et al., 2013). Assuming Yahoo was aware of the nature of the attack on LinkedIn and took immediate action to protect its users, it seems to have taken the company at least two years to switch one small algorithm for another, illustrating the challenges involved in such an operation and perhaps suggesting difficulties with the organisation’s security planning and maybe even with its overall software architecture.

Aside from the vulnerabilities of weak encryption, there exist vulnerabilities in the implementation of stronger cryptographic protocols. Much attention was drawn to the Alibaba Group’s UC browser in 2015 following the leaking of classified documents by a former NSA contractor Edward Snowden. The leak suggested that unencrypted geolocation and other data obtained from the browser were used to track its users. Following the revelation, the Citizen Lab, a research laboratory based in the University of Toronto, carried out an independent study on the UC browser which showed that user privacy was compromised; however, the lab could not confirm if the weaknesses that they found were those that were highlighted by the leak. In 2015, the UC browser had approximately 500 million users, most of whom were located in China and India. A later study on another browser popular in China, Tencent’s QQ mobile browser (Knockel et al., 2018),

25 The attacker precalculates hashes of passwords before the attack and simply compares hashes found in the attack with those precalculated hashes.
26 These large numbers indicate that some users set up numerous accounts and some were likely to have been fraudulent accounts.
27 See, for example, The Guardian (2017).
showed that data belonging to its hundreds of millions of users were also vulnerable to a so called man-in-the-middle attack with ‘state actor capabilities’. This research is of particular interest as it demonstrates weakness in ‘textbook RSA’ implementations. Such implementations are considered to be poor cryptography but are nonetheless in use and provide the only security option freely available to many millions of Internet users. Whether these weaknesses are accidental or by design is open to debate – the Chinese companies involved both conduct cutting-edge research, including research into quantum computation, meaning there is strong support for the ‘by design’ argument in the revelations made by Snowden. If the content of these documents is accurate, it is likely that no browser is secure. However, security issues also exist elsewhere on the Internet, including in areas where surveillance is unlikely to be currently a contributory factor.

The ‘Internet of Things’, for example, makes extensive use of Radio-Frequency IDentification or RFID tags. These tags or motes work wirelessly and remotely and carry only around 2,000 bytes making it impossible for them to support strong cryptographic protocols. Technologies in this early stage industry are still in a phase of intensive evolution and despite guidance from organisations such as the European Union Agency for Network and Information Security (ENISA), for example, there is as yet no clear policy for security of the Internet of Things. This lack of policy, coupled with its generally weak security, creates a potential point of vulnerability where the Internet of Things’ cyber-physical systems join the Internet proper (Shah et al., 2016).

Even in the absence of quantum technologies, the security of systems which interface with the Internet has been demonstrated to have considerable vulnerability. However, there are key areas which appear, at least, to be considerably more secure and resilient. Areas such as banking and finance in general, utilities such as national electricity grids and water and sensitive industries and governments are generally better protected than social media...

28 As RSA is described in textbooks with no enhancements.
29 A mote or remote is a wireless transceiver that also acts as a remote sensor.
30 There exist a small number of industry specific IoT security frameworks and best practice guidelines, all of which are still in the in development phase. There exists no overarching standard to date. See, for example (Microsoft, 2018).
31 Non Internet facing systems are also at security risk, but attacks on these systems require onsite access.
platforms, for example. Significantly though, any attacks on these critical systems are likely to have serious and far reaching consequences.

From their early design phase, security in critical systems must be carefully planned and managed according to relevant industry standards and organisational policy. This is in stark contrast to the often ad hoc arrangements of less critical systems. However, as critical systems evolve to meet changing requirements, or in response to security threats and attacks, weaknesses will appear in their security. If properly managed, weakness can generally be detected and analysed and appropriate modifications made to the system to reestablish and maintain the desired level of security. Ideally, this cycle continues until such a point as it is decided that, for security, cost or operational reasons, the system should be replaced. However, this cycle can be broken and consequently vulnerabilities in security may appear. Sometimes, large complex systems may be insufficiently understood by those who manage them and as the systems grow in complexity through maintenance and modification, understanding lessens and vulnerability increases. Although software companies may issue advice on operating system and networking security etc., it may be difficult for organisations to interpret and incorporate different strands of security information into a coherent secure policy for their organisation, or the recommended security measures may simply be beyond budgets. This is especially true in areas and times of political instability or economic challenge, rendering systems that are critical to infrastructure vulnerable to attack.

Since 2014, the computer systems of Ukraine’s state bodies, infrastructure, media, transport and politics have repeatedly been the target of cyber attacks. Russia has been widely accused of backing the attackers but denies any involvement. The scale of the attacks is unprecedented with, for example, more than 6,500 attacks on state institutions over a two month period in late 2016 alone. These attacks exploited a wide range of security loopholes and ranged from a highly orchestrated operation in which the electricity supply from three separate substations was cut off in a single attack to attacks on Ukraine’s financial and transport sectors. It has been reported that the Sandworm group was responsible

---

32 Speaking to Wired magazine, the NATO ambassador with responsibility for cybersecurity commented ‘You can’t really find a space in Ukraine where there hasn’t been an attack’ (Greenberg A. 2017a).

33 Power was later manually switched on again by the electricity company’s engineers.
for at least some of the attacks which involved malicious softwares including BlackEnergy 3 and KillDisk (Fireeye, 2018). Sandworm specialises in trojan attacks and is believed to have targeted ICS/SCADA and energy companies worldwide; it is one of several Advanced Persistent Threat (APT) groups currently operating globally (Greenberg, 2017). The US government has reported finding BlackEnergy malware on the networks of American power and water utilities, although here security was adequate to prevent damage (Greenberg, 2017a).

It has been suggested that the cyberwar on Ukraine has served as a de facto training ground for groups such as Sandworm and other APT groups with some attacks first seen in Ukraine quickly appearing in other jurisdictions. Believed to have originated in Ukraine, the NotPetya malware was responsible for a global rapid cyber attack in June 2017. The malware obtains user credentials from an infected host and uses them to connect to other points on the network, thus propagating the malware. In this way, just one machine infected with the malicious software can infect an entire system. NotPeyta, ostensibly a ransomware, has a highly unsophisticated ransom collection mechanism but considerable data destruction and encryption capabilities and consequently is considered not to be a true ransomware but rather to be designed to cause maximum disruption and financial loss to its targets (LogRhythm, 2017). It is likely that this malware was used as a test or reconnaissance attack. NotPeyta appeared just one month after WannaCry, another rapid cyber attack malware which caused major disruption in Spain, the UK, Russia, Japan, France and Taiwan. Believed to have originated in North Korea, this ransomware counted the British National Health Service (NHS) and Spain’s telecoms company, Telefonica, among its victims. WannaCry\(^{34}\) exploited a weakness in Microsoft’s Windows operating system for which a security patch had existed for about one month before the attack (Mathews, 2017a), highlighting the delay some critical service providers have in implementing security updates.

**The roles of artificial intelligence in security**

AI has recently entered the field of cyber security, with companies offering machine learning (ML) based defences against some of the most difficult and

---

\(^{34}\) One month before the WannaCry attack, a group called The Shadow Brokers released details of the weakness that the ransomware exploited in Microsoft’s Operating Systems, it is alleged that the weakness was originally discovered by the NSA. See, for example: (Gibbs, 2017)
pervasive cyber attacks. A relatively sophisticated example comes from the UK company Darktrace\textsuperscript{35} which has developed an algorithm, Enterprise Immune System, that is capable of detecting and defending against malicious network activity in near real time through the use of unsupervised ML techniques. This type of machine learning allows the algorithm to detect known and novel threats by actively self-learning patterns of normal and abnormal network behaviours rather than depending on known rules, models or datasets. Darktrace’s software has been demonstrated, for example, to detect a new strain of ransomware in a network and to have the ability to counter that attack in a time frame of under one minute\textsuperscript{36}. The algorithm has also been demonstrated to limit an ‘exfiltration of data by an insider’ attack (theft and export of data to the Cloud, for example) (Viega, 2018). In principle, this self-learning approach provides an added layer of security by constantly searching networks and interconnected networks for anomalous areas in large data sets and making decisions to act when deemed necessary. In contrast, traditional approaches depend on searching for evidence that exactly matches pre-described attacks and so novel attacks and approaches can go undetected.

In cyber security, artificial intelligence is dual use: it has the potential to be used in both defence and attack. AI network security algorithms may be vulnerable to data poisoning attacks, for example, in which misleading data is introduced by an attacker. Such an attack might be used as part of a scheme to train a network to tolerate intrusion. It is also likely that unsupervised ML might be used in more sophisticated and labour intensive attacks such as spear phishing\textsuperscript{37}, for example, where AI simulates more human-like behaviours and so attacks more readily escape detection. The potential of this so called ‘adversarial AI’ is not fully known; however, attacks as diverse as speech synthesis for impersonation, attacks that subvert cyber physical systems such as self-drive cars and the automation of techniques involved in surveillance, for example, are expected in the near to medium term\textsuperscript{38} (Brundage and Avin, 2018).

\textsuperscript{35} According to its website (www.darktrace.com), Darktrace was founded in Cambridge, UK, in 2013 by mathematicians and machine learning specialists from the University of Cambridge, together with world-leading intelligence experts from MI5 and GCHQ.
\textsuperscript{36} Attackers often spend months inside a network before being detected.
\textsuperscript{37} Spear phishing involves an attempt to steal sensitive information from \textit{targeted} individuals via electronic means.
\textsuperscript{38} Here, near to medium term is within the next five years.
The advent of AI as a security threat poses enormous challenges, challenges that translate into increased financial burden for organisations with data to protect. It is likely that as AI matures there will be a cycle of rapid growth in both AI defence and attack technologies. In response to this perceived threat, in February 2018, a group of 26 specialists from a wide range of disciplines and institutions including Oxford University’s Future of Humanity Institute, Cambridge University’s Centre for the Study of Existential Risk, OpenAI and the Center for a New American Security\(^{39}\) published a report on the potential security risks of AI. The one hundred page document, *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (Brundage and Avin 2018), outlines the group’s understanding of current and potential future threats posed by AI to security and makes a number of recommendations for future research and development, as well as highlighting the importance of governance and policy. The report details what its authors see as risks related to the publishing of potentially sensitive details of AI research (a practice common to all areas of computer science) and strongly recommends collaboration between the research community and governments in efforts to anticipate and mitigate AI attacks. The report focuses on the need to develop policies and regulation that are informed by technology expertise and are properly enforceable in the domain without unnecessarily restricting research. The authors draw attention to examples of introductory resources for policymakers in AI (CNAS, 2017; Buchanan and Taylor, 2017).

Already, there is much ongoing research in the area of Adversarial AI in particular (Brundage and Avin, 2018) and digital security in general, for example, by the National Cyber Security Centre as part of GCHQ in the UK (National Cyber Security Centre, 2018). A high proportion of large scale projects include workshopping and similar initiatives involving representatives of large technology companies, government agencies and research institutions. Many of these projects focus on US security concerns and are often aligned with political research centres such as the Harvard Kennedy Belfer Centre for Science and International Affairs, which in 2018 lists three ongoing projects: Managing the Atom, a project concerned with nuclear non-proliferation and disarmament; Managing the Microbe, concerned with the threat of biological weapons; and the Cyber Security Project, which

---

\(^{39}\) The Center for a New American Security is a US-based bipartisan national security think-tank.
concerns itself with how cybersecurity will shape international conflict (Belfer Center, 2018).

The implications of maturing quantum technologies

Due to the fundamental vulnerabilities inherent in current AI systems (vulnerabilities to data poisoning and model inversion\footnote{In model inversion, the training data of a classifier is manipulated.} for example), an increase in adversarial AI attacks is likely as AI technologies become more pervasive. These attacks are expected to be especially effective, finely targeted and difficult to attribute (Brundage and Avin 2018). The task of mitigating an increasingly enhanced attack load, including diverse and dynamic threats coming from adversarial AI, will eventually be beyond the capabilities and budgets of many organisations and infrastructure agencies. One possible response to such challenge is to move to fortified secure platforms where a collective security is provided by an overarching well-resourced body. This is not an entirely new idea: already, for many organisations, the increasing complexity of the task of providing a secure integrated internet environment has been met, for example, by the use of Google Cloud\footnote{Machine learning tools and APIs; the enterprise Maps APIs; and also the Android phones, tablets and Chromebooks that access the cloud.} together with G Suite (a platform that provides secure integrated electronic mailing (gmail); document Cloud storage (Google Docs); device management (Google Mobile) and other related services) (G Suite, 2018).

As technology companies such as Google position themselves to provide more comprehensive secure options to businesses and infrastructure facing growing security threats to their operation, the slow response to recommendations of collaboration between the research community, industry and governments in addressing the growing menace of adversarial AI is likely to be too little too late. If technology companies can provide protection to businesses, infrastructure and possibly even some governments, that exceeds any other option that these organisations have available to them, it is likely that many will migrate their operations to these secure environments. It is probable that there will be at least some competition in the space; however, it is also likely that only a few small to medium sized organisations will be sufficiently resourced and competent to provide for an independent fully integrated secure environment as the adversarial
AI threat grows. Unless governments provide an alternative, small to medium sized organisations will have little option outside of the secure spaces provided by the technology companies.

In the near term, there remain non-Google alternatives for individuals and organisations requiring secure Cloud and communications facilities. Google is not the largest of cloud-based services: Amazon, Microsoft Azure and IBM are all technically bigger, as are China’s Tencent and Alibaba. Google, however, offers perhaps the most obviously and fully integrated secure platform, with an emphasis on seamless integration of services and machine learning. Significantly, Google also has a post-quantum cryptography programme and recently the company substituted the RSA algorithm in Google Chrome with New Hope, a post-quantum algorithm (Pascaline, 2016). The New Hope algorithm is known to be less than secure against certain post-quantum attacks – it has known vulnerabilities to attack by a system with quantum capabilities (Malloy and Hollenbeck, 2016). Consequently, it is unlikely that Google considers the New Hope algorithm to be a credible candidate for the replacement of RSA in Chrome; in its current iteration at least. It is more likely that the purpose of this exercise was to assess the company’s capability of swapping out encryption algorithms without any downtime or incident. If this is the case, the exercise was likely considered successful. In the absence of truly secure post-quantum algorithms, it is essential that such swaps be easily and immediately achievable at the first sign of attack in order to limit data exposure or loss and it is likely that the large technology companies are continually researching and assessing new candidate post-quantum cryptographic algorithms. As discussed earlier, NIST expects that post-quantum encryption algorithms may differ significantly in their underlying mathematics and design from those in current use, making swapping out encryption algorithms in a live system a truly challenging task, far beyond the capabilities of all but a very few organisations.

As we approach the post-quantum horizon, the time when a device can efficiently and cost effectively run Shor’s algorithm, it is possible that large, well designed and competently managed systems of commerce and infrastructure with good AI-enhanced security will remain adequately secure, even in the face of adversarial AI. However, every instance of RSA-based cryptography will remain potentially vulnerable to attack, as any development of quantum and hybrid quantum devices
for malicious purposes will naturally be covert in nature. For all but the most expert technology companies, preemptive substitution of current public key cryptographic algorithms with post-quantum alternatives brings a high risk of weakening security. As quantum technologies mature, the only remaining viable option for even large mature organisations and many governments may be to move to a commercial platform offering post-quantum secured, AI-enhanced cloud, device management and communication services. Such a scenario would afford Google – and any other company which emerges with similar capabilities – potentially enormous political leverage as that horizon comes into view.

The recommendation from the Brundage and Avin (2018) report that governments should partner with research institutes to address these threats is important. These partnerships should work toward the development of secure platforms in the public domain outside of commercial technology companies such as Google. These post-quantum platforms need only be relatively rudimentary; however, it is essential that they offer the required security at low or zero cost to the user. There is a role for the EU here and at some of the funds already allocated to quantum technologies could possibly be directed to such a project. For poorer countries, such development should be taken on by the open-source community, in coordination with NGOs and multilateral institutions. This is essential to protect against these countries becoming unduly dependent on technology companies for their infrastructural and national security.

Secondly, during the phase of development of post-quantum algorithms, exposure to and understanding of the quantum paradigm needs to be increased. As highlighted by Harrow (2012), in addition to using computers to solve problems, we also think in ways that are informed by the programming and use of computers in the current paradigm. To a large extent, we frame and attempt to understand and solve many of the problems we encounter in terms of ideas and methods of information transmission, optimization and error correction. Concepts as diverse as sentiment analysis, weather forecasting and cognitive processes are all described within the bounds of our classical computational paradigm and our understandings are thus limited by its constraints.

Ideas of security, cryptography and the nature of information itself are also subject to the paradigm in which they are conceived and operate. Quantum computation, like the quantum theory that lies behind it, is counter-intuitive when viewed
through a classical lens. It is impossible to anticipate the novel approaches that will emerge from the field of quantum computation unless we are engaged in the paradigm. It is essential that we democratise the understanding of quantum information theory and normalise the use of its concepts in order to ensure that the paradigm shift in computing is not the preserve of a small corporate elite and a few, often corporate sponsored, research institutions. The Microsoft corporation has already made available a high-level accessible independent development environment (IDE), the Microsoft Quantum Development Kit, which works with Microsoft Visual Studio (Microsoft, 2018a). More similar initiatives, ideally from the open source community, would improve popular exposure to quantum computation. Explicit coding skills may not be required to familiarize the public with quantum information theory; initiatives such as the development of games that operate in the quantum paradigm would also provide an attractive introduction to the field and should be sponsored by government funds. Only widespread popular adoption and understanding of the quantum paradigm can prevent undesirable monopolies.

The final strand of defence of an independent Internet in an AI-enhanced post-quantum era is the prompt introduction of appropriate and effective legislation. Such legislation should be developed in partnership with domain experts in information security, government policy experts and the research community to ensure that any new legislation can be implemented in such a way that its intention is properly realisable. This was not the case in the drafting of the European Union’s recent General Data Protection Regulation (GDPR) legislation. It has been reported that this legislation has been difficult to implement in many cases due to its lack of compatibility with the nature and workings of the domain in which it is intended to operate and in particular in the context of current machine-learning algorithms (Goodman and Flaxman, 2016). Furthermore, the way we make and implement legislation needs to be reconsidered here and in relation to information technologies in general. The legal processes we use must be fit for purpose and capable of anticipating change. This does not mean that legislation needs to predict the precise changes that will occur – this is not possible – but rather legal processes must be such that they are capable of responding to a dynamic system of shifting and interacting paradigms; this will require both new processes and interdisciplinary expertise. Scholars from the social sciences and humanities can and must engage with developments that have the most profound
implications for the future of human communications: mapping the potentially extraordinary computational horizon is far too important to be left to computer scientists – or, simply, to Google.

References


