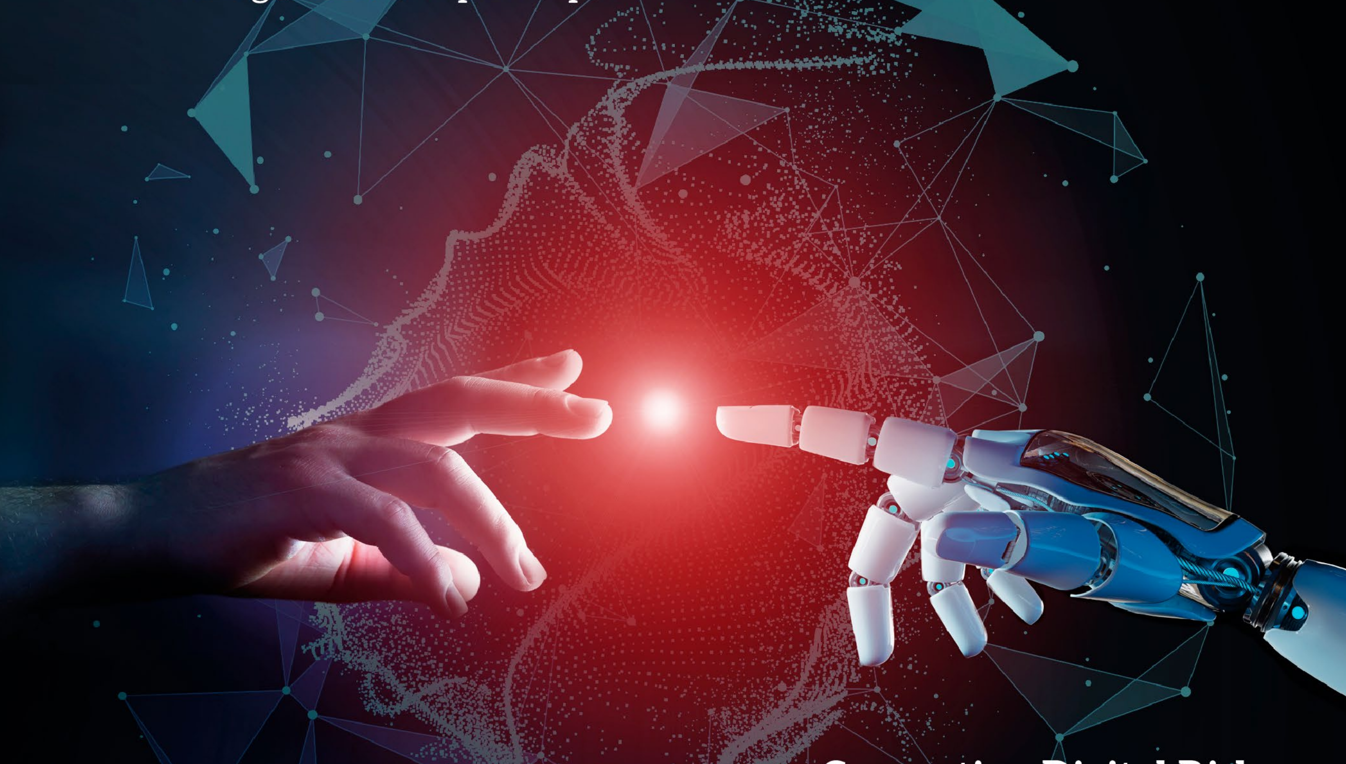
VOLUME 7 (2021) ISSUE 1

European Cybersecurity Journal

Strategic perspectives on cybersecurity management and public policies



Converting Digital Risk into Resilience in the COVID-19 Era

What Clinical Trials Can Teach Us About the Development of More Resilient AI for Cybersecurity

Interview with Lieutenant General Michael Vetter

ANALYSES • POLICY REVIEWS • OPINIONS

The Kosciuszko Institute

European Cybersecurity Journal

Strategic perspectives on cybersecurity management and public policies

The European Cybersecurity Journal (ECJ) is a specialised publication devoted to cybersecurity. The main goal of the Journal is to provide concrete policy recommendations for European decision-makers and raise awareness on both issues and problem-solving instruments.

ISSN: 2450-21113

Citations: This journal should be cited as follows:
”European Cybersecurity Journal”
Volume 7 (2021) Issue 1, page reference

The Kosciuszko Institute

Published by:
The Kosciuszko Institute
ul. Feldmana 4/9-10
31-130 Kraków

Phone: 00 48 12 632 97 24
E-mail: editor@cybersecforum.eu

Disclaimer: The views expressed in articles are the authors' and not necessarily those of the Kosciuszko Institute. Authors may have consulting or other business relationships with the companies they discuss.

Contents

Editorial

Barbara Wilk

Chief Editor of the European
Cybersecuirty Journal

Dear Readers,

It is fair to state that cybersecurity has never meant so much and has never affected as many people as it does today. Unprecedented acceleration of the digital transformation caused by the COVID-19 pandemic pushed us into a new realm, in which we are still learning how to function. Governments, companies, organisations, employers, and employees across the sectors needed to adapt to changes that nobody predicted. This black swan event affected every aspect of our lives and fuelled fears of what the future will look like.

It is not just the digital transformation that has accelerated – we also saw an increase in cyberthreats targeting companies, organisations, individuals, as well as critical areas of the functioning of economies, such as the health or energy sectors (of which a recent example is the May 2021 ransomware attack on Colonial Pipeline).

The multiplicity of topics that the reader will find in this issue – ranging from political and strategic matters, such as European digital sovereignty or cybersecurity policy development in the EU, to the increasing role of data, AI resilience, and importance of education – is an excellent read for the times when the digital dimension is entering the next level.
We have no choice but to understand these changes, and I am pleased that this journal contributes to raising awareness and spreading knowledge on the most pressing cyber topics of these days. We must learn how to function in a new reality which besides uncertainty also carries nearly limitless potential, and is powering our economy.

Foreword

Ciaran Martin

Professor of Practice,
Blavatnik School of Government,
University of Oxford

It is an honour to be asked to join the Editorial Board of the European Cybersecurity Journal and to be asked to write the foreword for this latest issue. The Kosciuszko Institute, CYBERSEC Forum and the Journal are among the most important international ventures in helping our continent, with our partners in other free and democratic societies across the world, deal with some of the biggest challenges in technological security of our time. It is a privilege to serve alongside such impressive and impactful people too.

This edition of the Journal brings to life some of the most pressing themes in cyber security. In recent months we have seen the scourge of ransomware emerge from behind the veil of corporate extortion, impacting the lives of ordinary people through the severe disruption of fuel supplies in the United States and healthcare in Ireland. And ransomware reminds us, perhaps more starkly than other forms of cyber intrusion, that online security in organisations is basically about risk management. So the paper from Olaf Schulz and Julia Jasinska explores those concepts of cyber risk and trust and how they apply to and within organisations. Robert Muggah's contribution focuses on resilience, the now essential concept in helping organisations withstand intrusions, wherever they come from.

These themes help us think through protecting the technology we have. But protecting the technology we're going to have, and keeping it open, free and economically sustainable in the face of competition from China's authoritarian alternative, is another crucial strategic cyber security challenge. That's why a series of offerings on the cyber security of AI, as well as maintaining public confidence in our free and open Internet through transparency around issues like surveillance, should be of great interest in this issue. So too should more technical examinations of security of the Internet of Things, one of the most important opportunities to encourage the adoption of a new and more robust model of commercial cybersecurity.

As always, there is plenty in this edition of the Journal for both the general and specialist reader, and anyone who cares about the online security of European societies.

The Path Towards European Digital Sovereignty

Interview with Lieutenant General Michael Vetter, Director General and Chief Information Officer of the German Armed Forces

Digital sovereignty in the EU context has been outlined as five courses of action: increasing innovation and research capabilities, promoting digital competency amongst citizens, using trustworthy IT, building up key technologies, and maintaining core command and control. How do these align with the wider European values such as fairness, transparency, democracy, and sustainability? Also, in what ways are existing projects such as GAIA-X, CONCORDIA, the European Single Digital Market, the 3 Seas Digital Highway contributing to the EU's digital sovereignty, as well as the digital sovereignty of our economies and societies?

First of all, I would like to thank you for having this important conversation with me today. The Bundeswehr also believes that the five areas you mentioned are essential building blocks for digital sovereignty. We are pleased that they have now been laid down in the Berlin Declaration on Digital Society and Value-Based Digital Government. This way, they serve as a common European foundation of our future digital policy.

All our efforts contribute to these lines of action. The CONCORDIA project is a good example. With this project, we are establishing and expanding European cooperation in the area of cybersecurity. The CODE Research Institute at the Bundeswehr University in Munich is acting as the project lead. The consortium is made up of 55 partners from 19 European nations. CONCORDIA will be a tool to strengthen research, market innovation, and capacity building and it will serve as a roadmap for cybersecurity research in Europe. Its vision is to create a community, build bridges, and lay the foundations for sustainable and close cooperation amongst all those involved. This is how we put European values into practice.

> CONCORDIA will be a tool to strengthen research, market innovation, and capacity building and it will serve as a roadmap for cybersecurity research in Europe.

Digital sovereignty entails that the EU takes control of its digital future by pursuing its own interests through cooperation and partnerships with

like-minded states, and Germany is perceived as one of the leading actors in laying the groundwork for that. In what ways does the EU seek to foster such collaborative efforts in mutually beneficial ways? What opportunities are there for other member states to take on greater roles as we shape the EU's digital domains and technologies that work for people?

Let me highlight that digital sovereignty is not about autarky or the implementation of protectionist measures. Rather, being digitally sovereign means to make sovereign and confident decisions on where we want to be independent and choose our own European way. For example, during Germany's Presidency of the Council of the EU, the EU member states agreed on rules for third-state participation in PESCO. This will benefit the collective security in Europe. More specifically, there are currently eight PESCO projects on Cyber/C4ISR (Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance). We initiated the PESCO project "Cyber and Information Domain Coordination Centre (CIDCC)" to improve the exchange of information on cyber issues with regard to planning, command, and control of EU missions and operations. Of course, these projects are open to the participation of other PESCO members and, with the new rules in place, also to third states.

The Bundeswehr and the German Ministry of Defence have undertaken numerous initiatives and considerable investment in research and innovation in a digital context and for cyberspace and cybersecurity. How can such capacities help bridge divides when engaging with stakeholders from the civilian sphere? What benefits do partnerships with civil society actors offer?

That is an important point. Currently, we are hardly able to solve the challenges in the cyber and information space on our own, which is something that will grow even more acute in the coming years. At the same time, I am convinced that we have excellent scientists, a strong industry and an innovative startup scene in Germany and in Europe that we can and must cooperate with. In recent years, we have launched a number of initiatives to strengthen and consolidate the ties between the stakeholders. We have an ongoing exchange with the industry in the form of regular strategic dialogues. Two key topics here are digital sovereignty and trusted IT. We are turning CODE, our own research institute, into a hub of federal IT security research and development. The Bundeswehr Cyber Innovation Hub (CIH) conducts targeted market analyses for us, establishes networks with the startup community, identifies new ideas, and validates existing solutions. Ground-breaking and visionary innovations in cybersecurity and the key technologies that are needed for this purpose will be introduced and promoted by the Agency for Innovation in Cybersecurity. To bolster our digital sovereignty, the recently established Bundeswehr Centre for Research on Digitalisation and Technology (dtec.bw) will provide a safe environment and foster research at both Bundeswehr universities. The centre will also feature an incubator to support potential startups. This will help turn ideas into marketable products and services, and keep intellectual property within the country. In addition, a diverse digital council supports us on our path towards the digital transformation of the Bundeswehr.

As you can see, a broad cooperation with many different stakeholders is essential for us. Through our innovation initiatives, we have created spaces for open and creative cooperation with businesses and the civil society. This is the only way to generate technological and digital innovations using new means and approaches and, in that way, also boost our organisation's agility. This will benefit both the Bundeswehr and the private sector.

> **Broad cooperation with many different stakeholders is essential for us. Through our innovation initiatives, we have created spaces for open and creative cooperation with businesses and the civil society.**

In your Keynote Speech at the CODE Annual Conference last November, you highlighted fragmentation across national boundaries

as a challenge to delivering on outcomes in the cybersecurity landscapes. In this, you underscored CONCORDIA's role and the establishment of the network of national coordination centres. What opportunities do you foresee for individual states to develop specialised national competencies or expertise in the digital context, especially for information and knowledge sharing to promote collaboration and EU-wide resilience? What other challenges and threats is the EU facing in pursuit of its objective to engage as a global digital player?

If we want to be able to cooperate, we first have to build our own competencies, which we can then bring into cooperative projects in a targeted manner. The same is true for the other member states. The European Cybersecurity Research and Competence Centre and the Network of National Coordination Centres will support our efforts by implementing targeted technology and research programmes in the area of cyber and IT at the European level, and by better coordinating research and knowledge across national borders.

To turn the EU into a global player in the digital sphere, we also have to reduce our technological dependencies. This includes completing the digital single market. The EU's Digital Markets Act and the Digital Services Act significantly contribute to this goal. We will support their implementation as a priority.

Generally speaking, we will all have to rethink our attitudes: from need-to-know to need-to-share, from isolated to connected, from silos to interconnectedness.

> To turn the EU into a global player in the digital sphere, we also have to reduce our technological dependencies. This includes completing the digital single market.

Emerging and disruptive technologies are one of the key challenges we are facing right now, raising calls for more regulation and accountability, especially when it comes to their military applications. To what extent do EDTs affect military defensive, offensive, and deterrent capabilities and strategies?

We are convinced that emerging and disruptive technologies generate opportunities. They will enable a better situational awareness, faster decision-making based on much more valid data and by this contribute to our security. More reliable data allows for making more objective, fully informed, smarter and therefore better decisions. Of course, technology always has to be used in line with our values and with international law.

However, EDTs also come with risks and challenges. These particularly concern our information security. One example is quantum computing and its possible ramifications for cryptography. These, too, are challenges that we have to address.

Non-state actors, such as entrepreneurs, technology companies, and NGOs, are taking on an increasing role on the global security stage, especially when it comes to advances in emerging and disruptive technologies. What incentives are there for private actors to partner with state actors (such as governmental bodies, militaries) and intergovernmental organisations (NATO)?

Just as we benefit massively from cooperating with non-state actors, so can they benefit from the public sector in several ways. Firstly, we are prepared to take chances with ambitious research projects and to provide funding for them. Secondly, we act as an intermediary between science, business and administration, by bringing together the know-how of universities, excellence in education, the ambition of the private sector and ourselves, the users. Thirdly, we create an environment at state universities that is conducive to the development of startups. This way, the latest research findings can evolve into marketable business models. Currently, we are implementing these ideas at our Bundeswehr universities via a "founders" initiative integrated in the aforementioned dtec.bw eco system.

So far, state cooperation with the industry and the economic sector has taken the form of a traditional client-contractor relationship. In recent years, however, it has become clear that further forms of cooperation are desired and required by both sides. Both sides should see themselves as partners and work closely together right from the start of a project or research scheme. Much has been achieved in this regard, for instance through a close cooperation on the maintenance and systems support of weapon systems. Another example is the newly published "Action Plan on synergies between civil, defence and space industries", in which the EU offers concrete collaboration incentives to the private sector and startups in particular. Furthermore, we are also breaking new ground in the area of cyber applications. Interested parties from the area of IT/tech can join our cyber reserve. By making available their skills to the Bundeswehr, they contribute to improving the cybersecurity of Germany. What is more, we have established partnerships in the area of cybersecurity that include the exchange of personnel. We need industry to participate in national security provision and to act as a long-term partner in maintaining our digital sovereignty. Our experience is that the industry itself is showing unprecedented interest in extended cooperation. It is not only economic aspects that play a role here. There is also great interest in doing one's part to support the public and the state.

Many of the emerging and disruptive technologies outlined in NATO's *Science & Technology Trends* report have dual use and serve both civilian and military purposes. To what extent does a blurring military-civilian divide change the strategic and operational security landscape?

In contrast to past technological developments, which were often initiated by the military, emerging and disruptive technologies originate in the private sector or the scientific community. This is why NATO is in regular and increasing exchange with the private sector, scientific institutions, public sector facilities and the civil society. The aim is for Allies to maintain their technological edge. The importance of close cooperation with the private sector and science requires innovative approaches to hiring specialists, screening investments and funding.

We know that as a client, we must become faster. To achieve this, we are, for example, reorganising the way in which we provide IT services. Standardisation, rapid scaling and top-down control are just some of the features of what we call our digitalisation platform. It will help us better keep up with the enormous speed of innovation cycles in the future.

In contrast to past technological developments, which were often initiated by the military, emerging and disruptive technologies originate in the private sector or the scientific community. This is why NATO is in regular and increasing exchange with the private sector, scientific institutions, public sector facilities and the civil society.

Autonomous capabilities are becoming increasingly embedded in a range of technologies, with both military and civilian applications. Some have voiced concerns over the moral, ethical, and legal implications as the human factor is argued to be further and further removed from the decision-making process. What is your take on this? Should we be afraid of autonomous and AI-powered systems?

Our position here is very clear. The Federal Government is opposed to weapon systems that take life-or-death decisions completely out of human hands. During the UN negotiations on lethal autonomous weapon systems (in short LAWS) in Geneva, Germany advocates for any future weapon system, including such with autonomous functions, to remain under human control.

We want to use new technologies within our clearly defined boundaries. By using AI, for instance, we aim at increasing precision, reducing complexity, and putting the operator or commander in a better position to take well-informed decisions. Using AI will enhance the protection of our personnel as well as civilians. Let me reiterate that this will always be done in line with international law and our Western values.

The choice, therefore, is not "AI *or* human?" We rather aim at effectively employing AI-based systems to support human decisions. Human decisions regarding the use of weapons remain indispensable, and so does human responsibility.

You also underscored in your Keynote, in the context of the COVID-19 pandemic, that we are witnessing additional challenges in the cybersphere in the form of disinformation, malware, fraud, theft, and other malicious cyber activity. How do we equip our citizens with the necessary digital literacy, knowledge and tools to combat or report these threats? Teaching digital literacy and technical skills in school settings might be one approach, but what about older populations or those more vulnerable to falling victim to cyber-attacks and crimes?

Digital literacy is inseparably linked to digital sovereignty. For this reason, it constitutes one of five fields of activity we defined. However, this is not only about individual military or civilian members of the FMoD's area of responsibility. Rather, sufficient digital skills must be consistently available across all command and leadership levels, both military and civilian. We must also get our industrial partners on board.

Since the area of responsibility of the FMoD represents only a fraction of society, I am glad, also as a citizen, that this important point is one of the key principles of the Berlin Declaration on Digital Society and will now be operationalised in the EU's Digital Compass.

In the context of COVID-19, it is imperative to designate and protect essential services, critical infrastructures, and supply chains from a host of threats, digital and otherwise. Does the EU's quest for digital sovereignty echo these needs? To what extent and in what ways?

The COVID-19 crisis has exposed a general need to quickly find innovative solutions and use digital technologies to maintain our ability to act. During Germany's Council Presidency, we at the Ministry of Defence therefore worked towards consolidating the EU's resilience in the digital realm and through digital solutions. What is more, the Programme for Germany's Presidency of the Council of the European Union named digital sovereignty as "a leitmotiv of the European digital policy". Digital sovereignty is a task that involves all parts of the state.

The Commission of the European Union, too, considers digital sovereignty to be a comprehensive challenge involving the society, the economy, and security policy. As early as 2016, for example, the Directive on the Security of Network and Information Systems created important prerequisites for strengthening the European ICT industry and for protecting critical infrastructures. The European Industrial Strategy published last year identifies key technologies in which to strengthen the technological and industrial base and to protect supply chains. The recently adopted EU Cybersecurity Strategy also reiterates the integrated approach to protect against various cyber threats.

What is your greatest concern or even fear regarding cyber and digital security trends we are observing right now? What action are you taking today to prevent or mitigate their effects?

I am concerned about our increasing technological dependence. That is why I believe we should focus on digital sovereignty, the necessary ability to act in and control the cyber and information domain. It must be our goal to be able to accomplish our constitutional tasks securely, independently and without unwanted third-party influence. I cannot achieve digital sovereignty on my own, and neither can the German Ministry of Defence. This is a task involving state actors, academia, industry as well as the civil society in a truly comprehensive approach.

Another aspect is the exponential rate of technological development in many areas. It is becoming increasingly difficult to predict the consequences of technology. What is more, the gap between our implementation times and technological development is becoming larger. But we are working on this as well; for one thing, we are steadily optimising our planning and procurement processes. And we are creating drivers of innovation such as the already mentioned Bundeswehr Cyber Innovation Hub, the Cyber Agency or dtec.bw.

As you can see, we are tackling both the opportunities and the challenges with equal determination. Ultimately, our goal is common digital sovereignty. ■

Questions by the Kosciuszko Institute

Lieutenant General Michael Vetter is the Director General for Cyber/IT and Chief Information Officer in the German Ministry of Defence in Berlin since April 2019.
Lieutenant General Vetter joined the Luftwaffe as an officer cadet in 1982 and served as a Supply Officer and Squadron Commander in a Tactical Reconnaissance Wing.
At staff level he served in various functions such as Assistant Chief of Staff Support in the Luftwaffe Air Transport Command and Branch Chief in the Luftwaffe Support Command. At MoD level he served as Assistant Branch Chief in the Air Staff, as Branch Chief Logistics Plans and Policy at the Joint Staff and as Assistant Chief of Staff for Support at the Joint Staff.
From 2003 until 2005 he was the Military Assistant to the Chief of the Air Staff.
Lieutenant General Vetter commanded Luftwaffe Maintenance Regiment 2 in Diepholz from 2005 until 2007 and became the Deputy Commander and Chief of Staff Bundeswehr Logistics Centre in 2009. From May 2013 until December 2013 he served as the Deputy Chief of Staff for Support in ISAF Regional Command North, Mazar-e-Sharif, Afghanistan.
From 2012 until 2017 he commanded the Bundeswehr Logistics Centre located in Wilhelmshaven.
In 2017 he became the first Vice Chief Bundeswehr Cyber and Information Domain Service (CIDS) and the Chief of Staff CIDS Headquarters.
Lieutenant General Vetter holds a Master's Degree in Economics from the Bundeswehr University Munich. He attended the German General Staff Officers Course at the Federal Armed Forces Command and General Staff Academy in Hamburg. He is a member of the Royal College of Defence Studies in London and a graduate from the Defence Resources Management Institute/United States Navy Postgraduate School in Monterey/California.
Lieutenant General Vetter is married to Yasmin. They have six children. His hobbies include cinema, reading and sports.

Cybersecurity – the Heart of the EU Security Strategy

Interview with Despina Spanou, Head of Cabinet of the Vice-President of the European Commission Margaritis Schinas

First of all, thank you very much for taking the time to give us this interview. Let's begin by explaining the new Cybersecurity Strategy's place within the broader regulatory landscape of the Union. The new Cybersecurity Strategy presented last December was introduced as a key component of Shaping Europe's Digital Future, the Recovery Plan for Europe and the EU Security Union Strategy. How do these strategies complement each other? How cybersecurity might be a stimulus in rebuilding a post-COVID-19 Europe?

I think that it makes absolute sense that cybersecurity is a key component of all these documents (and associated to even more than the ones that you mention). One of the priorities of the von der Leyen Commission was to take cybersecurity out of its digital and technology silo, and to upgrade it to a central place in the security ecosystem. There is no security without cybersecurity. This is why the Cybersecurity Strategy is a fundamental pillar of the Security Union. As this Commission has set the digital transformation of Europe as one of its top priorities, it is important to equip our digital economy and society with the safeguards our citizens need.

You also mentioned the Recovery Plan for Europe. Obviously, we won't be able to step up our security preparedness and resilience in a digital world if we do not invest in this. This is why cybersecurity also finds its place as a priority in the Recovery Plan. Member States have been encouraged to make full use of the EU Recovery and Resilience Facility to boost cybersecurity and match EU-level investment. For example, under the new EU Cybersecurity Strategy and new rules to make physical and digital critical entities more resilient, the Commission proposes to launch a network of Security Operations Centres (SOCs) across the EU to detect early signs of a cyberattack and to facilitate proactive action to boost our joint risk preparedness and response at national and EU level. Several Member States have already submitted their national recovery and resilience plans with concrete actions for developing national SOCs. The EU will top up these national investments with additional EU funding. The EU is committed to supporting the new EU Cybersecurity Strategy with an unprecedented level of investment in the EU's digital transition over the next seven years, through the next Multiannual Financial Framework for the EU budget, notably the Digital Europe Programme and Horizon Europe, as well as the Recovery Plan for Europe.

Furthermore, cybersecurity has become a strategic geopolitical factor. This is another reason why you will find it across the board in EU's policy documents and as part of this Commission's political agenda. For instance, cybersecurity is in the agenda for EU-US relations adopted upon the appointment of the Joe Biden administration in December 2020 (*Joint Communication: A new EU-US agenda for global change*).

> **There is no security without cybersecurity. This is why the Cybersecurity Strategy is a fundamental pillar of the Security Union. As this Commission has set the digital transformation of Europe as one of its top priorities, it is important to equip our digital economy and society with the safeguards our citizens need.**

The pandemic has shown our new vulnerabilities and the strategy introduces quite a few projects and initiatives to answer this challenge, such as the EU Joint Cyber Unit and the Cyber Shield. The latter, as the name suggests, might be exactly what we need to respond to the evolving threat landscape – a shield against adversaries. Could you explain in what ways is the Cyber Shield going to protect the EU citizens, institutions, and infrastructure? What will be the main responsibilities of the EU Joint Cyber Unit and how does it aim to boost the Union's overall level of preparedness and resilience?

During the COVID-19 pandemic, we indeed saw that cyberattacks increased, especially on some critical sectors of our economy and society such as healthcare. We saw recently the example of how a cyber-attack on a national health system can paralyze an EU Member State's health services at a time when their patients needed them most, cutting off access to patient records, forcing cancellations of medical appointments and delaying COVID-19 testing. On top of that, this stolen sensitive personal data could potentially end up leaked or sold in the dark web. Cyber-attacks are a serious threat to our infrastructure and society's well-being. This is why it is essential that we step up our capacity to work together in Europe to prevent and respond to cyber threats in a coordinated and effective way, to protect our internal market and the EU's citizens.

When it comes to security, EU has constantly worked to strengthen its security preparedness, focusing on making Europe more resilient to cyber-attacks, which could affect our digital and even physical life. The EU now needs to be agile and ensure it has the means to detect threats at early stage and deter these increasingly sophisticated and frequent cyber-attacks or cyber incidents. We can build on our culture of information sharing within the EU, and mobilise the full spectrum of legal and policy tools at our disposal to avoid and deter attack propagation across our internal market. We had already planted the seeds for this approach in the Blueprint on coordinated response to cyber-attacks that affect more than one European Union Member State, as well as in the Cybersecurity Act, which mandated the European Union Agency for Cybersecurity (ENISA) to work with the EU's Member States on their response in the case of a serious cybersecurity incident.

With the Cyber Shield concept, we are bringing everything together to achieve this EU-coordinated approach. The network of SOCs that I mentioned earlier is expected to improve EU's cyber resilience, through faster detection and response to cyber incidents, at national and EU level, and structured and coordinated operational cooperation as well as a mutual assistance mechanism in times of crisis in several ways. First, by establishing national, regional or sectoral SOCs serving private and/or public organisations with real-time monitoring and analysis of network traffic to detect malicious activities and information sharing agreements with public authorities. Second, by leveraging state of the art Artificial Intelligence (AI), machine learning techniques and computing power to improve the detection of malicious activities. And third, by enabling dynamic adaptation in changing threat landscape, the sharing of actionable cyber threat intelligence based on activities detected across borders, and notifying affected entities enabling them to take swifter action.

In addition, we are also trying to advance our knowledge on cyber threats, which are constantly evolving, and step up our operational capabilities. This requires high-level cybersecurity expertise, great coordination and operational maturity. This is the idea behind the Joint Cyber Unit, which will be an essential operational component of this cyber shield for EU.

As you mentioned, the threat landscape is expanding even further during the pandemic; the Commission is responding to that through the revised NIS Directive, also known as the NIS 2.0, and planning to address with this revision some of the limitations that were perhaps unforeseen as the original Directive entered into force in August 2016. What are some of the major shifts in the cybersecurity landscape you have noticed within the last five years that perhaps underscored the importance of adopting this revision?

The NIS Directive proposal was prepared as early as 2013, eight years ago. For the cybersecurity landscape, this period of time is an eternity. While the threat landscape has changed resulting in challenges to the security of the digital and physical world, important and positive changes have also occurred on the policy side. Back in 2013, the approach was different. Today the EU, including its Member States, opts for a coordinated approach to the cybersecurity and overall security of the internal market and our European way of life. At the time of its adoption, the NIS Directive was already an ambitious proposal, but now we need to improve it and take on board all the lessons learned from its implementation.

Our proposal for a revised NIS Directive, also known as NIS 2.0, might not seem ground-breaking but this is because it is a very targeted review. It is adapting some parts where we see the need for more harmonisation. For instance, the cybersecurity cooperation between the public and the private sector, and among industries, is still not mature across the EU. This has led to considerable disparities in terms of incident reporting. This culture of secrecy, of threat intelligence not being shared, has to be reversed, because it can cost society dearly. You are not protecting your organisation by not reporting an incident, which could potentially result in having whole system stall and whole parts of the society cease to operate – because here we are talking about critical sectors such as transport, energy, healthcare or banking. This is why we stepped up enforcement in our NIS2 proposal.

The second part where there was not a level playing field concerned the unclear delimitation of the scope of the NIS proposal. The absence of a harmonised designation of operators of essential services or critical infrastructure companies led to a very uneven situation across the Member States. This is why we decided to provide thresholds that would allow a more even spread of the identification of critical infrastructure services under the directive. When it comes to the sectors to be covered, we had to fill in certain gaps. It is inconceivable today not to include the pharmaceutical sector or the food sector along with healthcare, just as it is inconceivable not to cover public administration as critical infrastructure. One just has to look at the news in recent weeks and months on the hacking of the US Treasury and Commerce departments, on the impact on the US economy of the ransomware attack on the Colonial Pipeline, or on Ireland's Health Service Executive.

The changes we proposed through the review of the NIS Directive follow the philosophy of the NIS 1 Directive, and this is why we see the negotiations going rather smoothly with the Member States. What is new, however, is that by presenting together the revisions of the NIS Directive and of the Directive on critical entities resilience to align their scopes, we are breaking the silos of digital and physical security to ensure the same level of protection to the most vital parts of our societies.

> **You are not protecting your organisation by not reporting an incident, which could potentially result in having whole system stall and whole parts of the society cease to operate – because here we are talking about critical sectors such as transport, energy, healthcare or banking. This is why we stepped up enforcement in our NIS2 proposal.**

The revision might not be ground-breaking, but it does introduce further enforcement that will in the end provide better results, like you said. Under the new strategy, member states are also encouraged to complete the implementation of the EU 5G toolbox. What are the overall takeaways regarding countries' progress in implementing the toolbox of mitigating measures, for example, and what would be the next key objectives and actions of the Commission in this regard?

On the day of the adoption of the Cybersecurity package in December 2020, the Commission published a report on progress in implementing the EU 5G toolbox of mitigating measures. The review shows that most Member States have followed the approach proposed by the Commission. Since the previous progress report of July 2020, we see that they made progress in using the tools, assigning the operators the 5G services and in bringing the toolbox criteria into national processes. There are still some differences between Member States, with some of them being more advanced in certain areas than in others. However, overall, Member States seem to be keen to continue the EU coordinated approach on the cybersecurity of 5G networks. The Toolbox has been perceived as a useful instrument providing comprehensive guidance, based on risks and an objective methodology, to help them with their national processes. We also see that partners of the EU are starting to look at how we achieved the Toolbox approach in this matter, including for other areas and for other future technologies. It is vital to maintain this momentum and commitment on implementing the 5G Toolbox as an EU coordinated approach. In the EU Cybersecurity Strategy, we have detailed the way forward, with key objectives and concrete actions to continue coordinated work at EU level. These objectives are to ensure future convergence in risk mitigation approaches across the EU, support continuous exchange of knowledge and capacity building, and promote supply chain resilience and other EU strategic security objectives.

Another thing that perhaps might cause some, let's say, barriers or blockades is the digital talent gap in both the private and public sectors and more broadly gaps in digital skills among European citizens. In the past you have highlighted that missing skills are a threat on their own and that one way to address the talent gap is by bringing in people from a variety of other sectors, not just the technology sector. Could you elaborate on what bringing people together from a variety of backgrounds could mean in practice, is it something that we do at the hiring stage or by forming coalitions between industries?

Indeed, solid digital skills are a cornerstone of a successful Europe's digital transition. Yet, the digital skills gap in Europe is a reality. That is why in the new Skills Agenda we adopted last summer, we have set for the first time a very high target for digital upskilling and reskilling: by 2025, 70% of the adults (or 230 million people) should have at least basic digital skills.

The cybersecurity skills gap is a key dimension of the digital skills issues, and it plays out at all levels. First, at the level of users, or consumers of digital products. People have now become very familiar with their data protection rights following the GDPR, which has been implemented in a way that encourages everyone to take care of their privacy online. However, we have not achieved the same level of familiarity yet when it comes to cybersecurity. Cyber hygiene is not something that is automatically followed by everyone who has a smartphone, a tablet or a computer. But it is precisely these everyday devices, and the lack of vigilance by their users, that open the door to malicious actors, who, without necessarily intending to harm us personally in the first place, use us as tools to create botnets, to serve organised crime, and to destabilise democratic and peaceful societies. This worrying trend can only grow exponentially in an increasingly digital society. This is why we need people who will bring cyber hygiene into general education as a skill that everybody must have if they are going to be owners of electronic equipment.

On the other hand, we are facing a massive shortage of cybersecurity experts, and especially in the public sector. There is a struggle between the public and private sector. The more attractive financial perspectives in the private sector make it difficult to attract and retain people with cybersecurity expertise in the public sector. But more generally, as you pointed out, there is a lack of enthusiasm from other disciplines to join the cybersecurity field, because it is perceived as a technical area, reserved only for ICT and digital experts. I am a strong advocate of bringing talents from everywhere into cybersecurity. Cybersecurity needs lawyers, policy-makers, managers, people who understand economics, geopolitics or psychology. It is a truly interdisciplinary field. This is why, in addition to strengthening the cybersecurity skills of ICT specialists, it is crucial to provide for training opportunities for the non-experts. We also need to encourage people at a younger age to enter the field, explaining that cybersecurity is something that gives you an assured job – and a job in a critical profession, which can have a huge societal benefit. Therefore, we need a bigger place for cybersecurity in the education system as a whole, as well as cybersecurity training for the public and private sectors. Public-private partnerships can help this to happen. Let me highlight that the cybersecurity skills gap is not an EU specificity: this is not an area where Europe is doing worse than the US – actually, quite the opposite.

It is therefore very important that we create hubs of knowledge and training. The Regulation establishing the European Cybersecurity Competence Centre, which has been adopted by the Council in April, should help to boost cybersecurity skills and expertise across the EU, by identifying where to bring together experts from the public, academic and the industry sectors, who have knowledge and who teach others, in an easy, accessible, non-costly way.

> **Cyber hygiene is not something that is automatically followed by everyone who has a smartphone, a tablet or a computer. But it is precisely these everyday devices, and the lack of vigilance by their users, that open the door to malicious actors, who, without necessarily intending to harm us personally in the first place, use us as tools to create botnets, to serve organised crime, and to destabilise democratic and peaceful societies.**

It's all interconnected. It seems like trust is the most fundamental value we should cherish to make all of this work and the EU also has that firmly enshrined in its approaches to cybersecurity policy, whether it manifests in communications that inform people of how their data is used as you mentioned with the GDPR or in the certification and accreditation requirements for technical goods and suppliers. The importance of trust cannot be ignored. How can we foster even greater trust, not just between member states but also between individuals? How the three areas of actions proposed in the new cybersecurity strategy can perhaps enhance this process?

I believe that trust can be achieved by giving people reliable assurances that what they use is safe. EU policy makers and authorities have a responsibility to ensure that products and services available in the EU's internal market are safe. In Europe, we do it for the physical safety of the products and services. We also ensure that a product recalled in one part of Europe is recalled in all 27 Member States. We have to move forward and follow the same approach for connected products, not just in terms of their physical qualities and safety but also of their security. From homes and cars to medical products and services, our whole economy and society is connected. We can no longer afford not to have clear safety rules for the cybersecurity of connected products and digital services. This is one of the areas we announced in the Cybersecurity Strategy for further reflection. For this, we have to work with hardware and software manufacturers. We have to give users the assurance that their connected products or offered services in the EU are safe.

Part of this assurance is the certification schemes that will be adopted in the coming years. The EU is the first place in the world to have a certification possibility for ICT products, services and processes. For instance, the work launched on a cloud certification scheme is very important – this is where all our data is stored. Certification can allow manufacturers or providers of such products and services to offer the necessary assurance to users and consumers about their security.

To this can be added the more general imperative of transparency in cybersecurity matters – to create trust, we need to talk about cybersecurity, pointing out who is responsible for what, and each being accountable for what we do. We have to learn from the bad experiences. If you look at product safety for instance, the products that are being recalled today are not the same products that were being recalled five or ten years ago, because through the recalls and the continuous monitoring of the market, we have learnt and improved, and so has the market. We have to do the same with the security of connected products.

Words like cyber and digital are quite often used to describe what the next decade is going to look like. You could say perhaps it's a sign of the times, especially now when much of our lives moved online. What do you believe should be on our digital security radar, not only in terms of threats but also opportunities maybe for the next five to ten years?

Our priority should be to achieve a safe and open connected world, i.e. a world in which everybody feels safe to be connected. For this to happen, action is needed worldwide. The Internet became successful and evolved because was open and accessible to each and every of us around the globe. We need to protect this approach for an open and safe cyberspace.

In order to have a safe cyberspace, we need to ensure high level of cybersecurity as well as to invest in cybersecurity diplomacy, because not everybody considers security the same way across the world. Furthermore, we should not forget that in a connected world, when we discuss security or trade relations between nations, digital has to come in, because it is part and parcel of our life. This is where we need a paradigm shift of not talking digital as if it was something separate or irrelevant from the rest. It is part of all our aspects of our socioeconomic life.

The digital transformation started 30 years ago, and today it is an integral part of our life but we have to make sure that we have in place all the safeguards in terms of security and privacy that we normally have in our physical world. In the physical world, nobody can enter your home without asking you, nobody can take your information without asking you, you have in place the necessary legislation, processes and the right authorities to enforce the laws. All these have to exist also in the digital world. We need to break the silos between the physical and the digital world. This is what this decade should be about.

Then, we need a technological revolution, especially in Europe. On Artificial Intelligence (AI), we need a revolution in our way of thinking. We need to draw on all our technological capacity and build products and services that will serve our society. The pandemic clearly illustrated how AI can benefit our society. Without Artificial Intelligence, our response to the pandemic would have been much slower. We had for instance the opportunity to sequence the virus in a high-performance computer in Spain, which allowed us to understand fast what the virus was about. We need technology that matches the challenges of today and tomorrow. And we need to be more courageous with the use of technology in our daily life. In recent months, we have shown that we are capable of rethinking our education and work systems by harnessing our technological tools, whether it be through teleworking or distance learning. But to be more confident and less hesitant, we need safeguards. In Europe, these safeguards are found in security and in fundamental rights that define our European way of life.

Questions by the Kosciuszko Institute,
Conducted by Ewelina Kasprzyk

Despina Spanou is the Head of the Cabinet of the Vice-President of the European Commission overseeing the European Union's policies on security, migration and asylum, health, skills, education, culture and sports.

Her work on security consists in coordinating all areas under the heading of the EU Security Union, ranging from counter-terrorism, cybercrime, cybersecurity to hybrid threats. In this capacity, Ms Spanou coordinates the implementation of the EU Security Union Strategy 2020-2025, the first ever EU Strategy encompassing the whole spectrum of security work in the EU.

Previously, she was Director for Digital Society, Trust and Cybersecurity at the Directorate-General for Communications Network, Content and Technology (DG CONNECT) of the European Commission. In this capacity, Ms Spanou was responsible for the European Union's cybersecurity policy and law, digital privacy, connected cities and mobility, digital health, eGovernment and electronic identification. She was responsible for the implementation of the EU legislation on security of network and information systems (NIS Directive) and for the negotiations of the EU Cybersecurity Act as well as cyber dialogues with the EU's international partners. Ms Spanou has served as a member of the management board of ENISA, the EU agency for cybersecurity, and of the Steering Board of the Computer Emergency Response Team for the EU Institutions, bodies and agencies (CERT-EU). She is a founding member of the Women4Cyber initiative and advocate for the need for more cybersecurity experts in Europe.

Despina Spanou has also been Director for Consumer Affairs at the Directorate-General for Justice and Consumers (2013-2017) in charge of consumer policy, consumer and marketing law, redress and enforcement, and product safety. Prior to this position, she was Principal Adviser in the Directorate-General for Health and Consumers. She was the Deputy Head of Cabinet of the European Commissioner for Health and Consumers Mr Kyprianou (2004-2008) and of the Commissioner for Health Ms Vassiliou (2008-2010). Despina Spanou started her career at the European Commission at the Directorate General for Competition (in 2003). She had previously practiced European law with the Brussels branch of a US law firm.

Despina Spanou is a member of the Athens Bar Association and holds a Ph.D. in European law from the University of Cambridge. She is of Greek and Cypriot origin.

CYBERSEC

6th European Cybersecurity Forum CYBERSEC GLOBAL 2020 & 4th CYBERSEC Brussels Leaders' Foresight 2021

KEY TAKEAWAYS

> Together Against Adversarial Internet means in fact: together against an adversarial world.
>
> Izabela Albrycht, Chair, The Kosciuszko Institute; President, Organising Committee of the European Cybersecurity Forum – CYBERSEC

CONTENTS

OPENING REMARKS

STATE STREAM

DEFENCE STREAM

BUSINESS STRE

FUTURE STRE

SIDE SESSI

DISCLAIMER

CYBERSEC EUROPEAN CYBERSECURITY FORUM

TOGETHER AGAINST ADVERSARIAL INTERNET

RECOMMENDATIONS & KEY TAKEAWAYS

6th EUROPEAN CYBERSECURITY FORUM – CYBERSEC GLOBAL 28-30 SEPTEMBER 2020 & 4th CYBERSEC BRUSSELS LEADERS' FORESIGHT 18 MARCH 2021

#CSGlobal20

www.cybersecforum.eu

#CSBXL21

DOWNLOAD NOW!

OPINION

Don’t Destroy Security and Privacy to Save It

JAYA BALOO

VICE-CHAIR, QUANTUM FLAGSHIP INITIATIVE; CISO, AVAST SOFTWARE

ABSTRACT:

Current challenges to strong cryptography are not coming only from technology advances but from global policy-makers. Unable to strike the right balance between privacy and security, we are often left with neither. This is a grim echo of an earlier battle, namely the crypto wars of the 1990s. That battle was won by privacy advocates and cryptographers, and we need to take those lessons moving towards a post quantum secure future.

Keywords: encryption, privacy, quantum computing, crypto wars

Introduction

Likely you’ve seen in TV shows and movies a scene where police or intelligence agents are hot on the trail of a villain and are trying to get information, usually crucial to saving lives. But the information is “encrypted”, putting the investigation and lives at risk. No problem, some stereotypical “hacker” character (in a hoodie) says, we’ve got a “backdoor” and can “break” the encryption. Voilà, a few (unnaturally fast) keystrokes later they have the information they need, they catch the villain, save lives, happy ending, story over.

But we know life isn’t like TV shows or movies. “Backdoors” like that don’t exist in encryption today. And while underhand trickery might make for a good story, the reality is that if it were true, we would all be living in a world that’s not safer but much less safe.

In the United States, there's an absurd saying: "sometimes to save a village you have to burn it". The saying supposedly comes from the Vietnam War and points out the absurdity of using a noble end to justify horrific means to that end when those horrific means destroy the very intended noble end itself.

We're looking at the same absurd situation when it comes to encryption. People with noble ends – stopping crime, terrorism, child pornography, and human trafficking – are proposing means to those ends – weakening encryption with backdoors – that would in fact destroy the very end goal of security and privacy for everyone.

The situation is made even more complicated when we look at the very likely coming impact on current cryptography from quantum computing, which is looming as a force of its own that threatens past, current, and potentially future encryption and by extension security and privacy for everyone.

This creates a real-world ethical challenge for us in cybersecurity. How do we protect people's security and privacy while also trying to solve the pressing problems of stopping crime, terrorism, child pornography, trafficking – and staying mindful of the threats posed by the coming sea change quantum computing will likely bring?

Ethics is never easy: it's a discipline that requires grappling with questions and this is no exception. Part of the process of ethical grappling is understanding the full scale of the problem set and that's what I'm going to do in this essay, as well as highlight possible solutions that I feel best balance the right (or least wrong) answers to these problems.

How do we protect people's security and privacy while also trying to solve the valid problems of stopping crime, terrorism, child pornography, trafficking – and staying mindful of the threats posed by the coming sea change quantum computing will likely bring?

I believe these issues can be addressed in a balanced way. But like all good solutions, it requires several nuanced elements to come together. We should not weaken encryption, now or in the future. We should continue to work with legitimate agents that need legitimate access to information that may be encrypted to find effective, targeted ways to get that information without weakening encryption for everyone. And by continuing to support the strongest encryption possible, we can mitigate the risks that all encrypted information – past, present, and future – faces from quantum computing.

Why Weakening Encryption Is Actually Ineffective and Dangerous

We are today witnessing a revival of the "crypto wars" of the 1990s/2000s.

As a refresher, in that era governments, e.g. of the United States, treated encryption products like munitions with strict export controls around it. For example, you could use 128-bit encryption on your Windows NT 4.0 webserver in the United States, but if you wanted to stand up that same server in France, you'd have to use the weaker 40-bit encryption version. At the same time, law enforcement agencies pushed hard for a special backdoor for encryption that (in theory) only they could use. This all culminated in the failed "Clipper chip" initiative.

Those Crypto Wars ended with the Clipper chip dying and export controls being lifted. In the years since, we've seen global adoption of strong encryption in conjunction with the exponential growth of life on the Internet. Arguably, the stellar growth we've seen in the past 15 years or so is a positive result of how the Crypto Wars of that era played out.

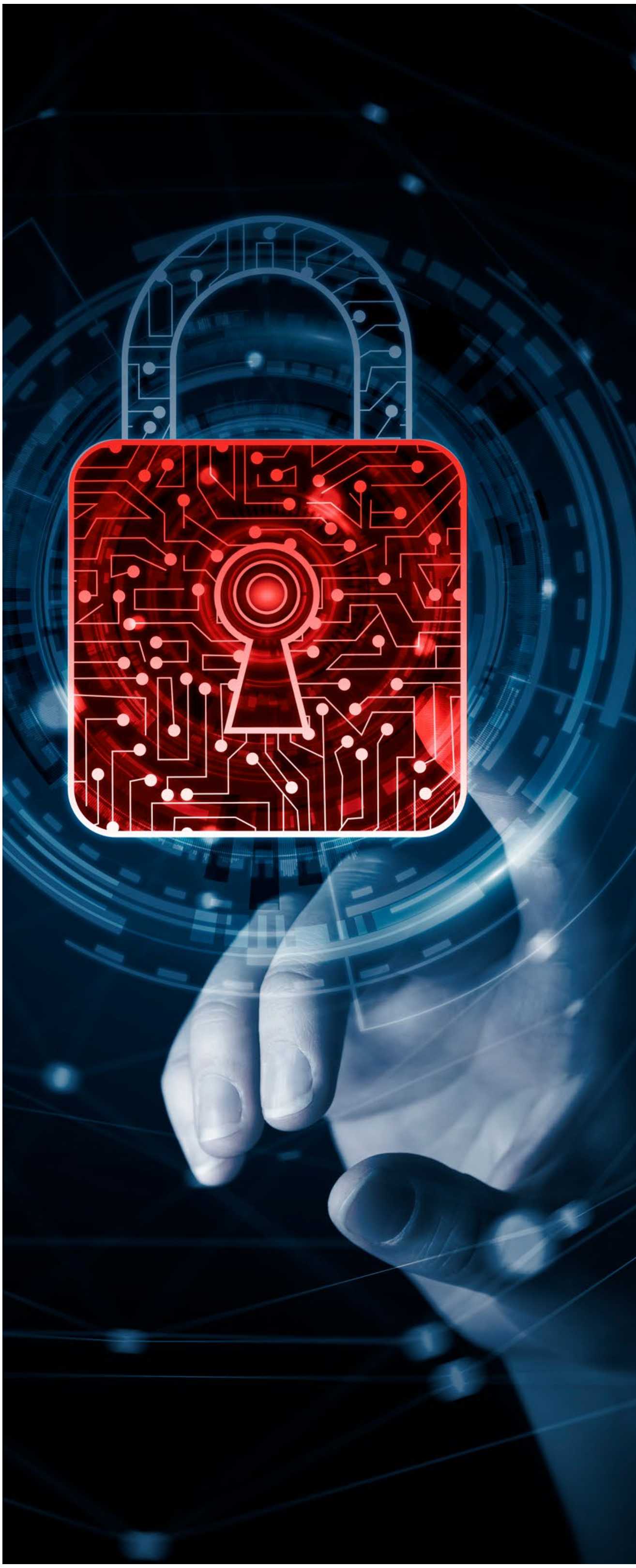
Today, however, the basic ideas behind the Crypto Wars are back on the table. After a series of high-profile events where law enforcement claimed that strong encryption was impeding their work, we saw law enforcement and governments once again calling for backdoors in encryption, ostensibly only for their use. The most notable example of this was in the wake of the December 2015 San Bernardino attack where law enforcement encountered problems accessing encrypted data on the attacker's iPhone and said that in this case a backdoor would be critical for them to have.

The arguments for backdoors being made today mirror those from the Crypto Wars. The objections to those arguments remain fundamentally the same. Implanting backdoors or weakening encryption in any way jeopardises security, privacy, and data integrity for everyone at all times.

Implanting backdoors or weakening encryption in any way jeopardises security, privacy, and data integrity for everyone at all times.

Benjamin Franklin famously said in *Poor Richard's Almanack* in 1735: "Three may keep a Secret, if two of them are dead". That saying holds true with encryption. The idea that special backdoor decryption keys will always ever only be used by trusted entities presumes perfect technical implementation and flawless process execution, neither of which are realistic. Aside from the risks of loss or misuse of the legitimate keys by legitimate entities, the reality is that the chances any built-in backdoor will be independently discoverable by other untrusted entities are very high.

Put simply, building any kind of backdoor into encryption means it's already broken on day one. It's only a question of when, not if, untrusted entities learn that it's broken and how to exploit that for their own gain. And once encryption is broken, it's broken for good. We've seen this already with encryption schemes that have had to be retired because the increase in computational power broke them. For example, the 40-bit encryption you had to use in the 1990s outside of the United States isn't used today because it's essentially useless.



Often discussions around the benefit of encryption backdoors to law enforcement rely on the analogy of a wiretap on a physical telephone line. Wiretapping with a court order is a legal activity with a long accepted history. However, this analogy is specious and inaccurate when we talk about encryption backdoors. Wiretapping is a practice that legally breaks the confidentiality of specific individuals at a specific time by targeting the relevant endpoints: it doesn't disrupt the confidentiality and integrity of all calls of everyone using the phone system at all times, which is precisely what encryption backdoors do.

It's also worth continuing the wiretapping analogy in a different direction to show that there are viable alternatives more in line with traditional wiretapping methodologies. Law enforcement and intelligence agencies have a number of alternative tools at their disposal that can gather information which is encrypted in transit without relying on backdoors and that target specific individuals the way wiretaps do. With appropriate warrants, these agencies can target individuals' computers and devices in a way that accesses data that has been encrypted in transit in an unencrypted form on the device. There are also ways to break encryption on specific devices without breaking encryption for everyone at all times.

There's also a practical point around effectiveness. If a specific encryption algorithm is known to be backdoored, the very people that this backdoor is meant to track will simply use other, non-backdoored encryption for their purposes.

All of this is to say that while there are reasonable arguments for the ends that advocates of backdoors want to achieve, the means to those ends are actually both ineffectual and dangerous, putting everyone's security and privacy at risk in a fruitless quest to try and (understandably and legitimately) target a small, admittedly dangerous subset of people.

How Quantum Computing Will Challenge Current (and Past) Encryption

We are in the middle of a developing revolution in computing that could be bigger than the computing revolution itself and possibly as big as the industrial revolution: quantum computing.

While we're not there yet, even some of the most conservative forecasters estimate we're only ten years away from quantum computing being a major disruptive force. And when we look at quantum computing, one of the major ways in which it will be highly disruptive is what it will mean for past, present, and future encryption.

We are in the middle of a developing revolution in computing that could be bigger than the computing revolution itself and possibly as big as the industrial revolution: quantum computing.

To understand the impact, it helps to cite an example from television. In the HBO series *Silicon Valley*, the main characters realise that they've created an artificial intelligence (AI) that is on the verge of being so powerful that it will be able to break all known encryption, easily, in a matter of days. They recognise that this literally puts the entire Internet at immediate risk because there will be no way for there to be secure payments or communications; basically, security or privacy as we know them will be gone.

I won't spoil the story and tell you how that ends, but the important thing to understand is that with quantum computing we really are facing that scenario in real life, and facing it in just a few years.

Fortunately there is work being done in encryption by both physicists and cryptographers on two separate solution tracks that can potentially meet this challenge. But the question remains whether it's going to come fast enough to prevent this kind of apocalyptic reckoning. A key point in the work of both these groups though is that we need to be strengthening current encryption today as much as possible as soon as possible to mitigate the potential negative future impact of quantum computing on encryption.

However, we don't have encryption in place today that we can confidently say will withstand the power of quantum computing. And we've already seen what happens when encryption algorithms that were once thought to be "secure" are broken by increased computing power.

The Data Encryption Standard (DES) was a 56-bit encryption algorithm developed in the 1970s. In 1999, nearly 30 years after it was devised, a DES key was broken by distributed.net and the Electronic Frontier Foundation in 22 hours and 15 minutes, effectively killing the algorithm. It's worth noting DES was 16 bits stronger than what people were limited to for export during the Crypto Wars, meaning that encryption was even weaker.

After DES was broken, people who built and used server software, browsers, websites, and mobile devices all had to go through a protracted, painful process of changing encryption algorithms to ensure they were all using more secure encryption. It was a process that took years. And any resting data coded with DES that remains accessible is now functionally not encrypted. Like in the *Silicon Valley* episode alluded to, information encrypted with DES effectively became plaintext overnight.

As quantum computers come into their own in the coming years, this is what we have to look forward to as more encryption algorithms that once took the lifetime of computing power to break are cracked in hours or even seconds.

How Moves to Weaken Encryption and Quantum Computing Collide

We stand at a point in time right now where people are lobbying to weaken encryption with backdoors while at the same time we can sense the near future when existing encryption algorithms without backdoors are likely to fall to the power of quantum computing.

These two trends run the risk of colliding and making the current situation around encryption and by extension security and privacy worse for everyone. And they portend to make the future even bleaker.

Current encryption is facing a clear and present danger already from quantum computing. Moves to weaken existing encryption with backdoors will only make the risks we're facing significantly graver. When something is already weak in the face of a coming onslaught, further weakening it, no matter how good the ends are, is the wrong and dangerous thing to do.

Further weakening existing encryption with backdoors will only hasten and make more likely the cryptographic apocalypse that *Silicon Valley* fictionally outlined and that we've seen hints of in the demise of past encryption algorithms.

Our focus around encryption now should be on strengthening, not undermining encryption, so as to better prepare for the impact of quantum computing.

And as I've outlined, the approach of introducing backdoors into encryption is reasonable but misguided: it sacrifices security and privacy for everyone in a quest to understandably target a small but dangerous subset of individuals.

Conclusion

So what should we do to address these two colliding streams today?

We should be doing all we can to strengthen encryption to mitigate the likely impact of quantum computing. As I said, there are promising trends in this space and they should be pursued with all due speed.

We should be doing all we can to strengthen encryption to mitigate the likely impact of quantum computing.

We should recognise the valid ends those arguing to weaken encryption with backdoors have but stand firm in saying that's a wrong and dangerous direction. We can work to find other means to those same ends, ones that protect security and privacy for everyone while giving governments and other legitimate entities the tools they need to meet those ends in an appropriately limited and targeted fashion, just like with traditional wiretaps.

It's not a simple answer: it's a nuanced answer. But that gets to the heart of ethical conundrums: there's almost never a simple answer, nor should there be.

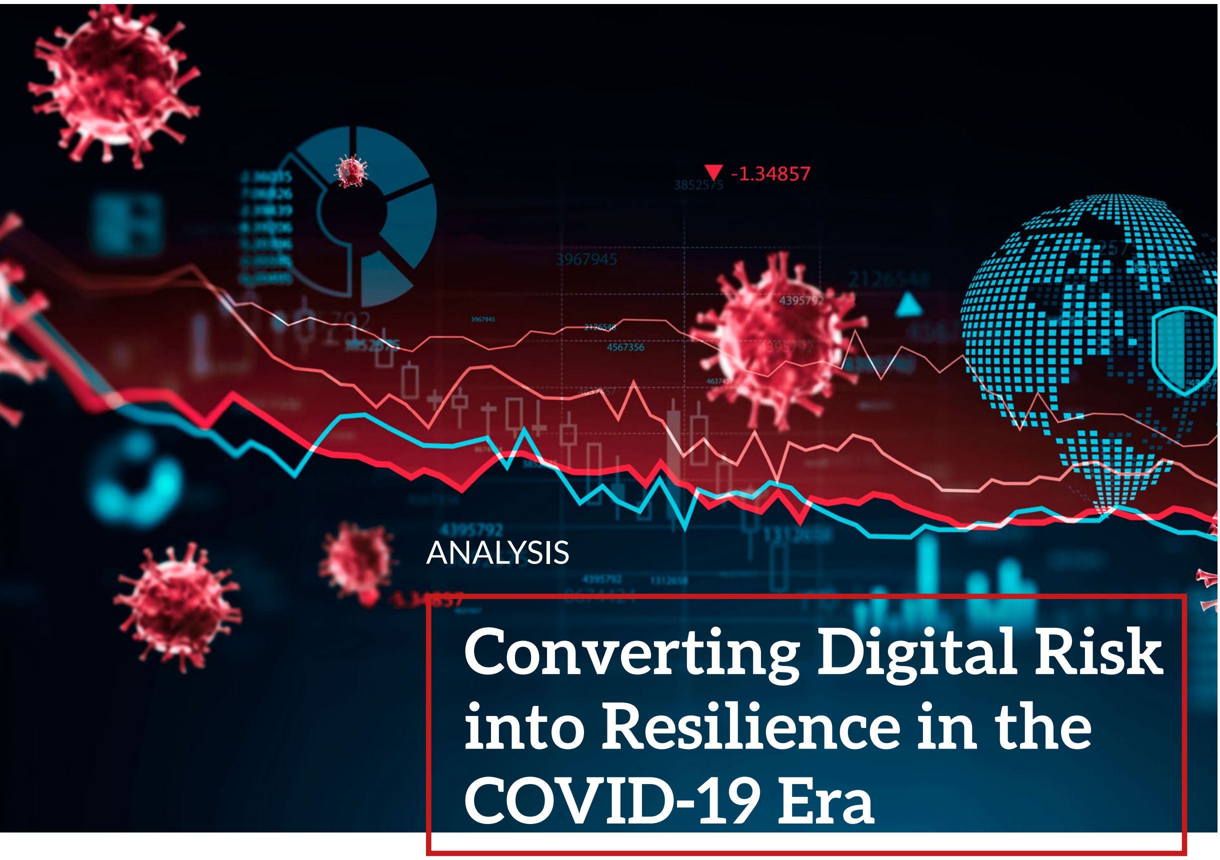
These issues reflect the fundamental tension between individual freedom, security, and privacy and the collective need for safety and security. Those values are almost always in tension but it's a healthy tension that reflects some of the best and most important characteristics of our open and free society. We want to preserve freedom, security, and privacy. And we don't want to destroy those very things in our efforts to preserve them. ■

About the author:

Jaya Baloo is Avast's Chief Information Security Officer (CISO) and joined Avast in October 2019. Previously, Ms. Baloo held the position of CISO at KPN, the largest telecommunications carrier in the Netherlands, where she established and lead its security team whose best practices in strategy and policy are today recognized as world leading. Prior to this, Ms. Baloo also held the position of Practice Lead Lawful Interception at Verizon, and worked at France Telecom as a Technical Security Specialist.

Ms. Baloo is formally recognized within the list of top 100 CISOs globally and ranks among the top 100 security influencers worldwide. In 2019, she was also selected as one of the fifty most inspiring women in the Netherlands by Inspiring Fifty, a non-profit aiming to raise diversity in technology by making female role models in technology more visible.

Ms. Baloo has been working in the field of information security, with a focus on secure network architecture, for over 20 years and sits on the advisory boards of the NL's National Cyber Security Centre, PQCrypto and Flagship Strategic. She serves on the audit committee of TIIN capital, a cybersecurity fund, and is also a member of the IT Committee of Sociale Verzekeringsbank. Ms. Baloo is currently a member of EU Quantum having been a member of the EU High Level Steering Committee for the FET Quantum Flagship from 2016 - 2017. Ms. Baloo has spoken widely at high profile conferences such as RSA, TEDx and Codemotion on topics including Lawful Interception, VoIP &amp; Mobile Security, Cryptography, and Quantum Communications Networks. Additionally, Ms. Baloo is a faculty member of the Singularity University since 2017, where she regularly lectures.



ANALYSIS

Converting Digital Risk into Resilience in the COVID-19 Era

ROBERT MUGGAH

PRINCIPAL, SECDEV GROUP; CO-FOUNDER, IGARAPÉ INSTITUTE

ABSTRACT:

The COVID-19 pandemic accelerated digital transformation around the world. The digitalization of government services, commercial transactions and social interactions generates benefits. But the digital transition also has a dark side, including sharpening digital divides and environmental costs. It is also giving rise to diverse online harms, from misinformation and disinformation to hate speech, extremist content and massive cybersecurity risks. Governments, businesses and civil societies must actively cultivate digital resilience - anticipating, adapting to, recovering and learning from digital threats. This requires approaching digital risk as an enterprise-wide concern; assessing and quantifying the scope of digital risk; ensuring that company leadership understands emerging risks; and developing a playbook to appraise and respond to digital risk.

Keywords: digital transformation, digitalization, cybersecurity, digital resilience

The digital revolution – a process that started in the early 1990s and is continuing into the present – is fundamentally changing the way we live and interact. The COVID-19 pandemic is accelerating the global shift toward digitalization while simultaneously widening digital divides and generating new digital vulnerabilities. Although the digital economy is speeding-up recovery and expanding opportunity, governments, businesses and civil society will need to navigate the digital transition with care. Those who successfully invest in inclusive and sustainable digital transformation, anticipate online risks and cultivate digital resilience will thrive. Doing so requires developing a mindset to adapt to and benefit from the upsides of digitization and minimize its many downsides.

Global digitalization was disrupted, and then accelerated, by COVID-19. In less than a year, the disease outbreak transformed the form and functions of government services, commerce and social interaction both online and offline. Future historians will likely look back on the post-2020 period in much the same way we currently do about the aftermath of the Second World War, another era of a profound disruption. That the COVID-19 pandemic occurred at a time of deep geopolitical divisions and heightened tension aggravated the crisis. The pandemic arrived at precisely the moment when international cooperation was most needed to deal with the myriad of global challenges from climate change to nuclear threats and cybersecurity (Muggah, Steven, and Torres, 2020). One thing is for certain: there is zero probability that we are returning to the 'old normal'.

The Bright Side of the Digital Revolution

Notwithstanding the many health-related, political, societal and economic traumas inflicted by the pandemic, one of its most dramatic effects was how it contributed to a surge in digitization. Around the world, the equivalent of a decade of digital onboarding occurred in less than ten months. Overnight, governments, businesses and citizens were forced to adapt to quarantines, restrictions and physical distancing. Entire workforces started working from home and moved their activities online. Countless supply chains were re-engineered, shifting from just-in-time to just-in-case. For many state officials, business people, and workers the choice was stark: go digital, or go dark.

Countless supply chains were re-engineered, shifting from just-in-time to just-in-case. For many state officials, business people, and workers the choice was stark: go digital, or go dark.

The COVID-19 pandemic precipitated a long-anticipated tipping-point in digital transformation. During the mid-1990s, technology enthusiasts predicted that the rapid spread of the Internet and powerful computing would generate new efficiencies, innovations and economies of scale. But the promised explosion of e-government and e-commerce never emerged after the dot-com bubble burst. But the digitization process has sped up since the 2000s. Over the past two decades, the world's digital footprint has increased exponentially. Presently, global IP traffic is almost 150,000 GB per second compared to 100 GB per day three decades ago. The spread of cloud computing, AI and billions of digitally connected devices set the stage for massive transformation of governance, commerce and everyday interaction.

Just before the arrival of COVID-19, the digital economy was growing more rapidly than the real economy. Depending on how it is defined, its total value amounted to roughly $11.5 trillion, or 15 percent of global GDP (The World Bank, 2016). Researchers believe this could rise to as high as $37 trillion, or 26 percent of GDP, by 2040 (Huawei & Oxford Economics, 2017). States with mature information and communications infrastructures – such as Estonia, Finland, Ireland, Singapore, South Korea and Taiwan – are well positioned to benefit from this growth. Advanced and emerging economies alike stand to gain if they can leverage new technologies to optimize processes and production, reduce transaction costs and digitize their supply chains. But progress will be more gradual until they overcome structural issues related to the generation, storage, processing and transfer of data.

COVID-19 has accelerated the digital transformation in clear and visible ways. One of these is near ubiquitous data and connectivity that is powering the new economy, with around 60 zettabytes that were produced in 2020 and almost three times as much expected by 2025. Dependence on cloud computing jumped by a third in 2020 (Kaur, 2021). Network operators registered as much as a 70-percent increase in the demand for Internet and mobile data services (Beech, 2020). Videoconferencing sky-rocketed by 700 percent last year (Sherman, 2020). Not surprisingly, the valuation of social media and remote conferencing companies soared. As impressive as these gains are, the digital transformation has not benefited everyone equally.

The Dark Side of Digitization

Although first movers are profiting, the digitization of government and commerce has failed to narrow the digital divide. Wealthy countries and companies are still far more digitally connected than poorer ones. It will be hard to close the gap. This is because success in the digital economy is determined not by the number of mobile phones and wireless connections, but by the ownership of infrastructure, code and data. Richer countries in North America, Western Europe and East Asia house well over 90 percent of the world's data centres, while Latin American and African states are home to less than 2 percent (Datacenters.com, n.d.). The US and China account for over 75 percent of cloud computing, 75 percent of all patents related to blockchain, and 50 percent of spending on IoT. Between them, they have over 90 percent of market capitalization in the world's largest digital platforms.

Although first movers are profiting, the digitization of government and commerce has failed to narrow the digital divide. Wealthy countries and companies are still far more digitally connected than poorer ones.

Some countries, companies and sectors are benefiting far more from digitization than others. A relatively small number of countries including the US (35 percent), China (13 percent), Japan (8 percent) and European Union states (25 percent) are reaping the benefits of the global digital economy (United Nations Conference on Trade and Development (UNCTAD), 2019). Likewise, a handful of firms – Amazon, Alphabet, Apple, Google, Facebook, and Microsoft alongside Alibaba, Baidu, Huawei, Tencent, WeChat and ZTE – have achieved dominant market positions and account for 90 percent of all revenue and profits (Oreskovic, 2019). Major retailers and manufacturers are restructuring and digitizing, or lowering the risk. Most businesses are going virtual in the hopes that they may benefit from network effects and greater competitiveness.

Making matters worse, the digital economy is generating serious negative externalities, including ratcheting-up climate change (Alorse, 2019). Despite the efforts of some tech firms to clean up their act, they are still considered among the most unsustainable and environmentally damaging in the world (Unwin, 2020). In order to meet voracious demand for hardware, they are ramping up extraction of rare earth minerals and other precious metals like cobalt (Watts, 2019). Technology redundancy and planned obsolescence are contributing to mountains of waste (Harris, 2019). Most worryingly, the expansion of Internet services is consuming about one-tenth of global electricity production (Clifford, 2019). The shift to cloud is scaling up energy consumption and carbon emissions, including from coal-fired power plants (Mills, 2013). The servers, cooling systems, storage drives and network devices of some of the world's largest data centres consume more than 100 MW of power, the equivalent of 80,000 US households (Energy Innovation, 2020). Bitcoin mining alone consumes over 7 GW, the equivalent of seven nuclear power plants and a carbon footprint comparable to that of New Zealand (Browne, 2021).

Meanwhile cybercrime, especially ransomware, has also increased exponentially. Digital attacks have impacted critical infrastructure, health facilities, federal and municipal government services, and corporations (Muggah & Goodman, 2019). With governments and companies becoming more dependent on complex Internet and cloud-enabled business models, they also expose themselves to more digital malfeasance. Over the next five years global companies risk losing an estimated $5.2 trillion in value creation due to cybersecurity attacks. And yet a minority of executives and insurers are confident in Internet security (Abosh & Bissell, 2019). Last year one of the most audacious cyber-attacks ever undertaken was discovered – a cascading supply chain attack – that could change the way the Internet is managed moving forward. No one is safe, and we have only seen the beginning (Muggah, 2021).

Digital attacks have impacted critical infrastructure, health facilities, federal and municipal government services, and corporations. (…) Over the next five years global companies risk losing an estimated $5.2 trillion in value creation due to cybersecurity attacks.

Managing Digital Risk

Some of the most successful governments and businesses are adept at turning risks into opportunities. Seizing digital opportunities begins with an understanding of the many dimensions of digital risk. It requires acknowledging the direct and indirect impacts of digital transformation –how technologies are changing and transforming government regulations, business efficiencies and client preferences – including the issues of privacy and data protection. At a minimum, companies need to understand and inventorize their total exposure. Digital risks necessarily impact on earnings and revenue. They are also intrinsically linked to technological choices and awareness of the regulatory environment. The truth is that one's reputation in the digital age is potentially global and instantaneous. It can also be destroyed in the blink of an eye. Companies need to set their risk tolerance. How much risk are they prepared to assume in order to achieve new economies of scale?

There are at least four straightforward principles to thinking about how to identify, mitigate and build resilience to digital risks.

The **first** is to approach digital risk as an enterprise-wide issue and not just an IT issue. Digital risk is a combination of people, processes and technologies. Determining what matters and what does not starts with a risk assessment to help identify the most valued assets, their location, means of protection, and access to them. It means deciding who is in charge and delegating authority and accountability as appropriate.

The **second principle** requires assessing and understanding the legal applications of digital risk. The regulatory environment for new technologies is fluid and fast-changing. It is shaped by politics from the international to the local levels. Concerns with foreign interference or privacy and loss of personal data are real and consequential. Corporations can be fined and executives can be jailed.

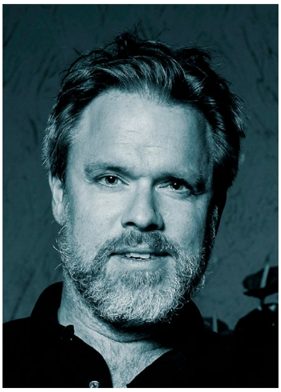
The **third principle** is to ensure that company leadership is on top of emerging risks and in constant contact with management. Executives must be able to answer the following questions: how secure are we and how do we know? What is the value at risk? What are the geopolitical and geo-digital threats to the company? What are the gaps? What do we need to know next? A constant dialogue with experts within the company and outside is more essential than ever in order to keeping a pulse on global trends.

The **fourth principle** involves setting up a clear playbook to appraise and respond to digital risk. Approaches will vary and evolve, but all companies need to start by quantifying the value at risk. This means assessing digital exposure as it relates to impacts on earnings, the amount of time required to respond to attacks, the capital and operational costs required, the loss of revenue, and the potential for fines. Firms should also create a risk register

– integrate digital threats into the business risk model – to easily communicate threats to corporate leadership. Risk management standards are key, as is applying them, so they provide the right metrics to drive decision-making.

The COVID-19 pandemic is forcing companies, governments and societies to address digital risks in a dramatic fashion. Even with the roll-out of vaccines, it is likely that many of these threats will endure for the foreseeable future. Business travel will not return at the scale of the past. Remote working will continue for many, especially as companies shutter their headquarters and move to more distributed work models (Statt, 2021). Cloud adoption will contribute to a wholesale redesign of enterprise networks. It is also intensifying concerns related to the protection of access to, and integrity of, data – a point made painfully clear with the Solar Winds hack (Muggah, 2021).

What makes the present moment exceptionally complex is that these dramatic changes are occurring during a period of intense geopolitical volatility. And while some level of uncertainty is inevitable, future pandemics, climate change and digital risks are not. There are no easy answers to how comprehensively protect ourselves from digital risks or build total digital resilience. But sticking to the fundamentals – and following key principles of digital risk management – will surely separate the companies that will thrive in the digital age from those that will not. ■



About the author:

Robert Muggah is a globally recognized specialist in cities, security and new technologies. He is a principal of the SecDev Group – a digital risk consultancy working at the interface of the digital economy and urbanization. At SecDev Group, Muggah helps city, corporate and non-profit leaders improve their future preparedness through high resolution data-driven diagnostics, strategy development, exponential leadership training, and public talks. In addition to his work at SecDev, he is also research director of the Igarapé Institute - a think and do tank - known for developing award-winning data visualizations and technology platforms to improve public safety.Muggah has spend decades tracking past and future trends in urban risk. He is faculty at Singularity University, senior adviser to McKinsey and Company, a fellow and adviser to the World Economic Forum, chair of the Global Parliament of Mayors, and a regular consult to the United Nations and World Bank. Muggah is the author of seven books and hundreds of peer-review and policy-oriented studies, including Impact: Maps to Navigate our Past and Future (Penguin, with Ian Goldin, out in 2020). His research is featured in global media, including the BBC, CNN, Economist, Financial Times, Guardian, New York Times, USA Today and Wired. He has given several TED talks viewed by millions, and speaks regularly at the Davos Summit. Muggah received his DPhil from the University of Oxford and his MPhil from the University of Sussex.

References

Abbosh, O. & Bissell, K. (2019). *Securing the Digital Economy: Reinventing the Internet for Trust. Accenture.* https://www.accenture.com/_acnmedia/Thought-Leadership-Assets/PDF/Accenture-Securing-the-Digital-Economy-Reinventing-the-Internet-for-Trust.pdf

Alorse, R., W. (2019, December 8). *The digital economy's environmental footprint is threatening the planet.* The Conversation. https://theconversation.com/the-digital-economys-environmental-footprint-is-threatening-the-planet-126636

Beech, M. (2020, March 25). *COVID-19 Pushes Up Internet Use 70% And Streaming More Than 12%, First Figures Reveal.* Forbes. https://www.forbes.com/sites/markbeech/2020/03/25/covid-19-pushes-up-internet-use-70-streaming-more-than-12-first-figures-reveal/?sh=32bbf48e3104

Browne, R. (2021, February 5). *Bitcoin's wild ride renews worries about its massive carbon footprint.* CNBC. https://www.cnbc.com/2021/02/05/bitcoin-btc-surge-renews-worries-about-its-massive-carbon-footprint.html

Clifford, B. (2019, April 8). *The alarming environmental impact of the Internet and how you can help.* Medium. https://medium.com/wedonthavetime/guest-blog-post-the-alarming-environmental-impact-of-the-internet-and-how-you-can-help-6ff892b8730d

Datacenters.Com. (n.d.). *Data Center Locations: Top Cities, States, Countries and Regions.* Retrieved April 6, 2021 from https://www.datacenters.com/locations

Energy Innovation. (2020, March 17). How Much Energy Do Data Centers Really Use? Energy Innovation. *Energy Innovation: Policy & Technology LLC.* https://energyinnovation.org/2020/03/17/how-much-energy-do-data-centers-really-use/

Harris, J. (2020, April 15). *Planned obsolescence: the outrage of our electronic waste mountain.* The Guardian. https://www.theguardian.com/technology/2020/apr/15/the-right-to-repair-planned-obsolescence-electronic-waste-mountain

Huawei & Oxford Economics. (2017). *Digital Spillover: Measuring the True Impact of the Digital Economy.* Huawei Technologies Co., Ltd. https://www.huawei.com/minisite/gci/en/digital-spillover/files/gci_digital_spillover.pdf

Kaur, D. (2021, February 3). *Cloud computing spend increased by a third in 2020.* Tech HQ. Retrieved from: https://techhq.com/2021/02/cloud-computing-spend-increased-by-a-third-in-2020/

Mills, M. (2013). *The Cloud Begins With Coal: Big Data, Big Networks, Big Infrastructure, and Big Coal: An Overview of the Electricity Used by the Global Digital Ecosystem.* Digital Power Group. https://www.tech-pundit.com/wp-content/uploads/2013/07/Cloud_Begins_With_Coal.pdf

Muggah, R. (2021, January 11). *Why The Latest Cyberattack Was Different.* Foreign Policy. https://foreignpolicy.com/2021/01/11/cyberattack-hackers-russia-svr-gru-solarwinds-virus-internet/

Muggah, R. & Goodman, M. (2019, September 30). *Cities are easy prey for cybercriminals. Here's how they can fight back.* World Economic Forum. Retrieved from: https://www.weforum.org/agenda/2019/09/our-cities-are-increasingly-vulnerable-to-cyberattacks-heres-how-they-can-fight-back/

Muggah, R., Steven, D., & Torres, L. (2020, April 23). *We urgently need major cooperation on global security in the COVID-19 era.* World Economic Forum. Retrieved from: https://www.weforum.org/agenda/2020/04/we-need-major-cooperation-on-global-security-in-the-covid-19-era/

Oreskovic, A. (2019, June 16). *This chart shows just how much Facebook, Google, and Amazon dominate the digital economy.* Business Insider. https://www.businessinsider.com/facebook-google-amazon-dominate-digital-economy-chart-2019-6?r=US&IR=T

Sherman, N. (2020, June 2). *Zoom sees sales boom amid pandemic.* BBC. https://www.bbc.co.uk/news/business-52884782

Statt, N. (2021, February 9). *Salesforce declares the 9-to-5 workday dead, will let some employees work remotely from now on.* The Verge, https://www.theverge.com/2021/2/9/22275304/salesfore-remote-work-9-to-5-workday-is-dead-flex-coronavirus

The World Bank. (2016, May). *World Development Report 2016: Digital Dividends* [World Development Report]. World Bank Group. https://www.worldbank.org/en/publication/wdr2016

United Nations Conference on Trade and Development (UNCTAD). (2019). *Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries* (UNCTAD/DER/2019). United Nations Publications. https://unctad.org/system/files/official-document/der2019_overview_en.pdf

Unwin, T. (2020, February 20). *Digital Technologies Are Part of the Climate Change Problem.* ICT Works. https://www.ictworks.org/digital-technologies-climate-change-problem/#.YBy1tuhKiUI

Watts, J. (2019, December 18). *How the race for cobalt risks turning it from miracle metal to deadly chemical.* The Guardian. https://www.theguardian.com/global-development/2019/dec/18/how-the-race-for-cobalt-risks-turning-it-from-miracle-metal-to-deadly-chemical

ANALYSIS

What Clinical Trials Can Teach Us About the Development of More Resilient AI for Cybersecurity

EDMON BEGOLI, PhD
AI SYSTEMS R&D SECTION HEAD, OAK RIDGE NATIONAL LABORATORY (ORNL), USA

ROBERT A. BRIDGES, PhD
CYBERSECURITY RESEARCH MATHEMATICIAN & CYBERSECURITY RESEARCH GROUP LEADER, OAK RIDGE NATIONAL LABORATORY (ORNL), USA

SEAN OESCH, PhD
CYBERSECURITY RESEARCH SCIENTIST, OAK RIDGE NATIONAL LABORATORY (ORNL), USA

KATHRYN E. KNIGHT
DATA ENGINEER, OAK RIDGE NATIONAL LABORATORY (ORNL), USA

ABSTRACT

Policy-mandated, rigorously administered scientific testing is needed to provide transparency into the efficacy of artificial intelligence-based (AI-based) cyber defence tools for consumers and to prioritize future research and development. In this article, we propose a model that is informed by our experience, urged forward by massive scale cyberattacks, and inspired by parallel developments in the biomedical field and the unprecedentedly fast development of new vaccines to combat global pathogens.

Keywords: artificial intelligence, resilient AI, cybersecurity policy, clinical trials, evaluation and test

Introduction

As human health is essential to a functioning society, vaccines have proven critically important for dampening and even thwarting many diseases that have plagued humanity (MacDonald et al., 2020). Yet, to be used, vaccines must be endorsed by regulatory bodies and are used only if they pass a set of rigorous, staged scientific trials aimed at (1) guaranteeing a large benefit-to-detriment ratio in use and (2) providing visibility into at least some scientifically grounded measures of efficacy for the public. By analogy, we argue that the world's economy, critical infrastructure, society, and culture depend on healthy, reliable, and functional networks, which in turn are reliant on computer network defence technologies (e.g., malware detection, network intrusion detection) as critical defences from cybercriminals, state-sponsored actors, and other adversarially minded groups. Yet, the scale and the size of the problem transcends any single tool or approach, requiring artificial intelligence-based (AI-based) approaches to help us cope with the scope, scale, complexity, and uncertainty of the problem. However, no policy for mandating minimal effectiveness, reliability, and transparency of these defensive, especially AI-based, measures exists. Enacting such a policy is a gargantuan undertaking, and we recognize that fact. In this paper, we take a proverbial bird's eye view of the well-established policies in critical domains, such as the vaccination development process, and zoom in to discuss organically occurring analogous testing of commercial off-the-shelf cyber tools currently underway. This provides a real-world example of how cyber analogues to clinical trial policies may strengthen cyber defence and structural efforts for future developments. The promises of AI and machine learning (ML) in computer defence, in particular, are considered.

We argue that the world's economy, critical infrastructure, society, and culture depend on healthy, reliable, and functional networks, which in turn are reliant on computer network defence technologies as critical defences from cybercriminals, state-sponsored actors, and other adversarially minded groups.

This discussion is particularly relevant in light of the recent SolarWinds attack (Tung, 2021), which was reported to have impacted over 250 US federal agencies and businesses (Schneier, 2021). In response to this attack, security experts have highlighted the need to "improve government software procurement" (Schneier, 2021). The authors' research organization, Oak Ridge National Laboratory, is currently aiding US federal agencies in testing intrusion detection technologies that are powered by AI/ML to ensure that the technologies selected are maximally effective (McDermott, 2020). Admittedly, the authors are not policy experts, nor do we have the necessary data to make detailed policy recommendations. However, we are reaching into our unique experiences with these novel technologies with the hope of informing future best practices and to spur conversations that can lead to the development of more effective cybersecurity policies for certification of AI-based cyber defences.

Background

In this section, we provide a brief background on the clinical trial process and the role that AI plays in cybersecurity. Then, we will make suggestions for how to map AI-based cyber tool development to the clinical trial process and exemplify how we are already using elements of the clinical trial process in practice to aid in the development of more reliable cyber tools. We will then discuss the implications of our experiences and what effective cybersecurity policy might look like in practice.

Rigor in Clinical Trials

To introduce a new vaccine that demonstrably immunizes against the target pathogen(s), the proposed treatment needs to undergo a sequence of rigorous clinical trials providing staged evaluations and gradual introduction of the new treatment. These trials, in general, consist of three or four stages (Evans, 2010). In the first stage, the proposed vaccine is evaluated for the proposed validity of its design; subject-matter experts evaluate the planned treatment to assess its merit. In the next stage (or a trial), the proposed vaccine is tested in a limited and controlled laboratory setting, often on lab animals. In the third stage, the proposed vaccine is tested with a limited patient population and in a double-blind protocol; that is, for an ideally large number of subjects, both the actual treatment and a placebo treatment are administered identically with both the recipients and those administering the treatment blind to who received real or placebo vaccines. Comparing the results of infection between the groups provides statistical tests to determine efficacy of the treatment in the real world. Once the efficacy and safety standards are reached, the treatment is approved for use in the general population, but the results of the application of the treatment in the population are monitored for safety, efficacy, and adverse effect, which could result in a recall of the vaccine. This is the fourth phase or stage of the clinical trial.

Artificial Intelligence Possibilities and Pitfalls

Two of the major concerns in network defence are (1) the ability to detect attackers once they are in the network and (2) the speed at which a response occurs once they are detected. AI-based security solutions have the potential to solve both detection and response problems. Examples in the cybersecurity market abound. Decision classifiers (used in supervised learning) are already replacing malware detection at the endpoint and network levels, while more sophisticated algorithm suites are packaged to identify suspicious behaviours via network traffic and other logged data and are meant to complement rule-based systems. Driven by insider threat concerns, a submarket of User and Entity Behaviour Analytics (UEBA) tools use ML-based models to detect anomalous user or machine behaviour in the network, which could significantly reduce the time it takes to detect attackers once they are in the network. Security, Orchestration, Automation, and Response (SOAR) solutions promise to increase automation of triage and response, thereby potentially reducing response time.

> **Two of the major concerns in network defence are (1) the ability to detect attackers once they are in the network and (2) the speed at which a response occurs once they are detected. AI-based security solutions have the potential to solve both detection and response problems.**

While AI offers unique capabilities that can help make computer networks more secure, it also increases the attack surface available to adversaries and presents new avenues for attackers to exploit (Liwel et al., 2019; Katzir & Elovici, 2018). More generally, there are numerous examples (Uesato, 2018) of AI vulnerabilities, in and outside of the cyber context. As recent research shows

(Li, 2018; Song, 2019), all of these qualities are also subject to adversarial exploitation and can be a source of privacy and security vulnerabilities. For example, the models can be trained on purposefully and intentionally tampered training data. This approach is known as data poisoning. The models trained on such "poisoned" data would in effect make AI models develop a "blind spot" for the phenomena that adversaries are interested in exploiting once the model is deployed in operations. Another approach (i.e. adversarial perturbations, data patches) manipulates the input to the already-trained model with an intention to confuse it. The goal of this kind of exploitation is to confuse the defences and have them mistake the malicious input for valid, therefore passing the defences undetected. In summary, the statistical machinery and computing power of AI can be and is used to poison, fool, or otherwise thwart a second AI system. This emerging, fast-growing research area, called "adversarial AI," is continually producing new methods. That dynamic makes the need for a continuously updated, rigorous, formal, and principled approach for testing and evaluating reliable and resilient AI all that more important.

One view (discussed informally (Gore, n.d.) and explored rigorously (Wang et al, 2007) is that security, physical or otherwise, is the practice of introducing asymmetry by creating enough difficulty that exploitation is no longer worth the cost. Because AI is providing powerful automation tools to either instantiate or alleviate difficulties, it stands to benefit from both defensive and offensive actors. While it is still too early to tell whether attackers or defenders will benefit most from AI, an argument can be made that AI will ultimately tip the balance in favour of the defenders if it significantly reduces the time required to detect intruders once they enter the network (Buchanan, 2020). This possibility is highly relevant to policymakers at the government level because the offense-defence balance impacts strategic stability, whether or not nation-states decide to go on the offensive (Jervis, 1978; Glaser, 1997). Overall, there is a need to make AI more resilient and reliable for it to be used in critical domains such as cyber defence.

While it is still too early to tell whether attackers or defenders will benefit most from AI, an argument can be made that AI will ultimately tip the balance in favour of the defenders if it significantly reduces the time required to detect intruders once they enter the network.

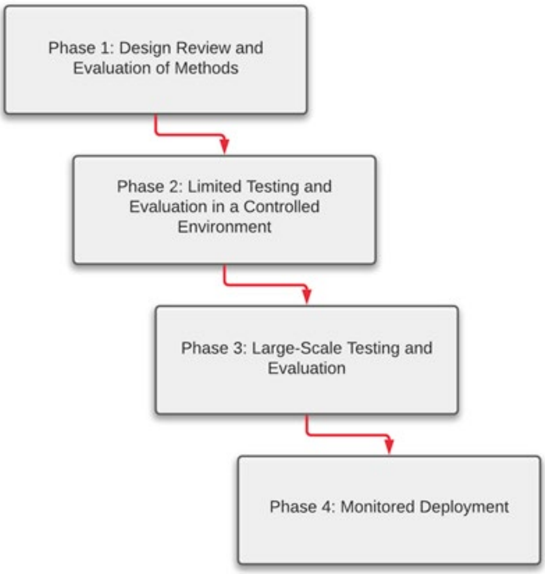
Taking Inspiration from the Structure of Clinical Trials

Here we consider the clinical trial stages and attempt a hypothetical mapping of their meta structure to the conceptual stages by which AI-based cyber defences can be rigorously and methodically evaluated for resilience to adversarial exploitation and for effectiveness in large-scale network defence. We argue that the challenges of defence-based AI research are analogous to the scientific challenges posed by medical treatment research, where a pathogen may be either known or novel, with treatments often requiring continuous (re) evaluation and improvement. As such, we propose four phases of evaluation comparable to clinical trials, with each phase examining the test design elements and methodology, approaches to controlled and large-scale evaluation, and ongoing monitoring requirements, respectively.

In Phase 1, the focus is on the rigorous evaluation of methods and design characteristics. The objective of this phase is to explore the theoretical and conceptual aspects of the approach before the larger-scale tests are conducted. The attempt is to answer the question "*Will it work?*" before further testing is attempted. Only small-scale tests are conducted to evaluate the critical design and methodological aspects of the approach, including its fundamental efficacy and the soundness of the defensive AI models.

In Phase 2, the next stage of expanded evaluation happens in a controlled environment and against a larger number of malicious and benign software components. Some network components can be simulated. The focus of this phase is on the effectiveness, reliability, and possible side effects of the approach, including critical vulnerabilities or inability of AI to respond to known malware, and the well-known adversarial AI vectors.

In Phase 3, the large-scale evaluation begins. It happens in an operationalized-like environment, with common network, system, and software components in place, and against a large library of malware (as well as "benignware"). The testing in this phase in all respects mimics a production-like environment, and it is designed to statistically match the population and the distribution of threats found in the "open". The emphasis of this phase is on the efficacy of the AI solution in the function of cyber defence and the success ratio against any known and speculative configuration of threats.[1]



Phase 1: Design Review and Evaluation of Methods

Phase 2: Limited Testing and Evaluation in a Controlled Environment

Phase 3: Large-Scale Testing and Evaluation

Phase 4: Monitored Deployment

Figure 1. A proposed structure of evaluation phases.

Phase 4 is a monitored deployment phase. The cyber defence solution is deployed to the production environment and operationalized. Following inspiration from clinical trials, the deployed defence is monitored continuously for safety and efficacy. Similar to the deployment of new therapies, the deployed AI-based tools can be recalled if they are observed to be vulnerable to adversarial exploitations, or if they do not function properly.

Emerging Lessons Derived from Practice

The recommendations presented in this work are inspired by two principal sources. One is our experience in working on a similar but less rigorously structured testing and evaluation program for AI-based cyber defence tools. The other inspiration is our experience in AI use in biomedicine and drug discovery. Naturally, the new inspiration occurred at the intersection of these two domains.

Related to AI-based cyber defences, we are engaged in an ongoing effort to strengthen the computer network defence tools of research sponsors in what is currently a three-part process. Each step requires an ideally impartial set of evaluators with technical expertise and cooperation of the companies developing and/or selling technologies of interest. (As it happened, we had devised the evaluation process that, as we observed later, happened to map to the first three stages of a vaccine clinical trial.)

First, subject-matter experts (e.g., account executives and engineers) from the companies or laboratories promoting a cyber-technology meet with an impartial third party (in this case, the research scientists) to convey and defend their technical approach. The goal is to allow the scientists to determine the merits based on the concepts put forth. Over time, this market research allows the scientists to see a wide variety of ideas and corresponding technologies in each subclass of the market (e.g., malware detection, UEBA, SOAR). Our experience with doing this type of market research with a few dozen companies is that while companies are generally proud to provide insights into the intellectual merits of their approach, the propensity for salesmanship clouds the merit of the discussion. This motivates the subsequent steps.

Second, an experiment is designed and carried out with a goal of deeply testing competing

[1] Threats such as complex, novel malware and advanced persistent attacks can be simulated using AI itself.

technologies head-to-head in a controlled laboratory environment. Medium-scale examples of such testing include tests as described in our concurrent work (Bridges et al.) where thousands of malware/benignware files of different types are used to evaluate endpoint and network-level detection abilities. Such a test provides great insight into consumers of those specific technologies and a glimpse into the state of the art. Examples of large-scale experiments in this vein include the AI ATAC Challenges (Naval Information Warfare Systems Command, US Department of Defense [NAVWARSYS], n.d.-a, n.d-b, n.d.-c). In the first, endpoint malware detection (commonly called antivirus, or AV) technologies were stress tested with 100K files in a completely automated, highly parallelized custom framework while many measurements (e.g., CPU, memory, bandwidth resource) were gathered. In the second, a full network with services (e.g., email, ssh), emulated and real Internet access, and emulated users, was designed and implemented to test network traffic analysers' ability to detect multistage attack campaigns. Hundreds of adversarial campaigns were waged against the network to identify strengths and weaknesses of the detection tools. (The third AI ATAC Challenge is still under development.)

These details are included to provide an idea of the scope of such experiments, and to show the development of the methods scientists are beginning to develop to measure how efficacious cyber defences actually are. Provided enough tools are included in these large-scale tests, the state of the art as used in practice emerges. For the first time, consumers can see gaps in defences across all choices, identify combinations of tools that maximally enhance defence (known as "defence in depth"), and researchers/technologists can prioritize next-step efforts. Importantly, policies mandating such testing will balance and scrutinize the marketing promises, often stated as a readily offered panacea for otherwise hard problems.

Step three, which entails a pilot study of tools in real-world conditions but with limited scope, has yet to be completed by our team. Such a study for detection and prevention tools would necessitate deploying each network defence tool in a network along with "honeypots"—computers donning realistic configurations but lacking in defence mechanisms—which would function as the analogous clinical trial "placebo". This study allows for direct comparison between control and experimental results (e.g., statistics can be computed for comparing the health of real network resources compared to their honeypot counterparts). Beyond tool efficacy, analyst usability experience measures should also be considered in this stage, providing a new dimension to this means of testing.

Because our previous experience is in testing security software, monitoring how tools perform when widely deployed in the real world is a natural extension of this. Yet, it is unclear how exactly to measure widespread performance, as this also may necessitate information sharing among disparate network operators. The aversion of network operators to share data notwithstanding, many modern security tools used in practice rely on cloud connections for updates and data processing, especially for applying AI models. This opportunity provides telematics feedback from each network across the globe to the vendor-operated cloud servers. The data informs threat intelligence and can strengthen AI models via retraining. Furthermore, the security as a service industry is growing, allowing providers to gather information from a wide variety of deployed security tools. In summary, private markets are leveraging opportunities to learn from widespread network defence tools in the wild, and the crowdsourced intelligence generated for consumers (e.g., network security professionals) is evidently worth the risk.

Notably, there are private and government efforts to catalogue software vulnerabilities, software platforms, types of weaknesses, and mitigations. This publicly available information has proven to be a great boon to the information security community and illustrates how all parties benefit from the upkeep and organization of information from those willing to contribute. Upon initial disclosure of a vulnerability, a structured entry appears in the Common Vulnerabilities and Enumerations (CVE), a publicly available index of software vulnerabilities hosted by MITRE, a private corporation. After appearing in the CVE dictionary, each vulnerability is ingested and enriched with much more information by the National Vulnerability Database (NVD), hosted and actively managed by the US National Institute for Standards and Technology. The NVD includes cross-references to other related information also held in enumerated standards (e.g., referencing the appropriate entry of the Common Weakness Enumeration (a dictionary of types of insecurities in software and hardware) and of affected software in the Common Platform Enumeration (a dictionary of software including vendors and versions). As the name suggests, CVE is the de facto standard for cataloguing vulnerabilities; for instance, security professionals, the NVD, and other databases refer to vulnerabilities by their CVE number. In addition to these government-assisted sources, many other public vulnerability databases exist (e.g., Levy, 2019). In parallel, as vulnerabilities and correlated information are being catalogued, vendors of the affected software are, in general, creating mitigations (e.g., software updates and patches), and posting in their public-facing bulletins, referencing the known vulnerabilities alongside their patches. Overall, through private and public efforts to continually catalogue, update, and make available well-organized and accurate information on security vulnerabilities and patches, a system and an infrastructure has emerged to quickly index and share security problems and their fixes. This provides a wealth of needed resources for security operations to protect networks.

There are private and government efforts to catalogue software vulnerabilities, software platforms, types of weaknesses, and mitigations. This publicly available information has proven to be a great boon to the information security community and illustrates how all parties benefit from the upkeep and organization of information from those willing to contribute.

Discussion

In an article discussing the regulation of AI-based systems that replace humans in decision-making processes, Mulligan et al. advocate "a move from a procurement mind-set to a policymaking mindset" (2019). While the motivation and application differ in cybersecurity, we also advocate a move from procuring the best existing AI-based defensive cyber-technologies to establishing policies that govern the entire life cycle of such technologies. These policies would be intended to establish accountability and transparency, which we believe would result in more effective AI-based tools for cybersecurity. As noted by Wieringa in a systematic survey of literature on algorithmic accountability, accountability "both exposes issues and promotes better behaviour because people know they are being watched" (2020).

In our own experience of testing ML-based tools for the US navy, we saw first-hand how rigorous testing and feedback can both enhance existing commercial tools (by providing feedback to the vendors) and aid in effective procurement. We suggest that the four-phase clinical trial process can be mapped and applied to the development of AI-based tools to provide both accountability and transparency. While we make no specific claims about the most effective way to implement this process via policy, in this section we discuss potential actions policymakers could take to implement each of these four phases and the limitations of our approach.

Phase 1

Phase 1 requires domain experts to evaluate a technology's methodology to determine soundness. While our experiences documented above discuss third-party consultants meeting with vendors for this purpose, a more scalable option is already in place for cybersecurity. Specifically, Gartner (Gartner for IT, n.d.) provides market research, in particular for cybersecurity submarkets, with a goal of providing information on the value of particular vendors' solutions. In addition, Mitre Att&ck Evaluations (The MITRE Corporation, n.d.) have entered the cybersecurity market research sector, with a goal of "enabling users to better understand and defend against known adversary behaviours through a transparent evaluation process and publicly available results." Both provide information on the pros and cons of cybersecurity technologies without rigorous testing. In short, industry is providing a Phase 1 analogue, and the challenge before us is to provide, either through policy or some other means, the credibility and integrity of these sources.

Phase 2

Multiple research efforts besides ours are creating benchmarking abilities (e.g., datasets, evaluation methodologies), and/or novel metrics and measures for AI and cybersecurity technologies (ATT&CK evaluations, 2021; Myneni, 2020; Ring, 2019; Shah, 2020), yet most such efforts conclude with an academic publication. Implementation of an AI and/or cyber-focused trials, inspired by actual clinical trials, will require strategies for determining which evaluation methods are merit-worthy, ideally automating them, and finally incorporating them into a regular process with publicly available results. What is needed are clearly defined and transparent criteria, approaches, and requirements for testing and evaluating the resilience and reliability of AI-based methods for cyber defence. The government can play the role of a trusted third party and an "honest broker" that can facilitate testing and evaluation of the AI-based cyber defence approaches supplied by commercial entities and private industry. In this role, the government can set standards for testing and evaluation, host "bake-off" events, and impartially track and evaluate results while remaining completely neutral and protecting the intellectual property of the participants.

> **Implementation of an AI and/or cyber-focused trials, inspired by actual clinical trials, will require strategies for determining which evaluation methods are merit-worthy, ideally automating them, and finally incorporating them into a regular process with publicly available results.**

As a thought experiment, suppose a national or international body curated, regularly updated, and publicly facilitated a test for some subset of cybersecurity tools, similar to Kaggle competitions.[2] The idea is that, for a specific class of cyber-technologies leveraging AI/ML, one can submit a tool and be provided with an evaluation automatically. Similar to R-value insulation ratings, food labels, and vaccine effectiveness statistics, transparency will benefit consumers, establish the state of the art, and even possibly promote healthy competition for vendors. It is critical that the testing methodology maintains both impartiality and integrity. Thus, it is imperative that third-party expertise remains a requirement for both administering and changing tests. It is also worth noting that deprecated test datasets may be made public for cybersecurity researchers. Finally, novel and varied tests may comprise adversarial AI techniques so as to provide a stress test of cyber tool resilience to these techniques. Overall, we believe such testing infrastructure, shared datasets, and the resulting transparency will enhance the development of tools in addition to providing needed information to the public.

Notably, existing initiatives can help evaluate the security and viability of contemporary AI models. One such initiative is the GARD program, orchestrated by the Defense Advanced Research Programs Agency (DARPA). The program is intended to create a platform for the evaluation of attacks and defences against adversarial AI. This initiative presents one possible way for purposeful and principled evaluation of AI models against adversarial exploitation. Furthermore, it is a model where the government serves as a host and convener of the approaches developed and executed by the academic and commercial entities.

[2] Kaggle designs and hosts ML competitions often by clearly and technically defining a desired task, making public a training dataset, and holding secret a testing dataset for evaluation of submitted technologies.

Phase 3

Similar to vaccine trials, pilot testing of AI and/or cybersecurity defence tools in a real network can provide insights into their effectiveness. For the analogue of placebos, we propose honeypots—computer systems donning no defences and possibly with known vulnerabilities present—with the goal of monitoring adversarial techniques. Ideally, honeypots will be instantiated alongside close-to-identical workstations and servers, the latter employing defensive measures under test. Overtime, monitoring of defended attacks on the actual workstations versus the undefended activity on the honeypots will provide measures of the efficacy of defensive techniques. This whole process can be expedited by red team events, where hired offensive computing professionals wage cyberattacks to identify weaknesses in the defences of the network and/or confuse or bypass AI models. Research for creating, tuning, and operationalizing such a process would be the next step to pursue such evaluations.

Phase 4

In the final phase, deployed tool monitoring provides a means of both recalling ineffective tools as well as retraining models to address attack vectors on the underlying algorithms as they arise. Having a public database of exploits against different classes of algorithms and methods for protecting an AI-based system against these exploits, including a way to access data sets to harden models if required, would be a good place to start. This database could also include attacks against specific tools, whether or not the tool vendor has provided sufficient protection against the attack, the release number in which the patch occurred, and a recommended course of action for organizations considering using that tool on their network. The CVE and NVD databases are used to track software vulnerabilities and offer a model for what this process might look like in practice. In addition, MITRE has a well-defined process for how to report such vulnerabilities (Researcher Reservation Guidelines, 2020). Because vulnerabilities in AI systems are fundamentally different from software vulnerabilities, the structure and content of such a database would look differently, but a similar reporting process could be used.

To ensure that AI-based tools are properly secured in operational environments, it would also be helpful to establish clear policies that security operations centres (SOCs) can follow to ensure a secure deployment as well as a pathway allowing SOCs to provide feedback that can be used to improve the tools themselves. Examples of helpful policies include guidelines for protecting training data that could be used or poisoned by an adversary and steps for integrating these tools effectively into the existing infrastructure. Regarding feedback from SOCs, Engstrom and Ho (2020) propose that having random subsampling of cases quality checked by a human operator will provide a more tractable measure of how an automated mechanism is performing in practice. Given that SOCs often perform manual investigation of logs and incidents, they are ideally situated to provide this type of feedback. Findings of investigations are usually well documented by ticketing systems and preserved by saving correlated network data in the tickets. Research and mechanisms for "closing the loop", (i.e., leveraging these ongoing manual efforts to systematically find and fix problems with automated tools) is burgeoning (e.g., see Veeramachaneni et al. [2016]), but no widespread practices have been established. Creating a clear pipeline to receive feedback on existing deployments from SOCs, addressing any identified issues, and then pushing resultant updates out to all SOCs using that tool would be extremely beneficial.

To ensure that AI-based tools are properly secured in operational environments, it would also be helpful to establish clear policies that security operations centers (SOCs) can follow to ensure a secure deployment as well as a pathway allowing SOCs to provide feedback that can be used to improve the tools themselves.

Recalling products, while used in many other sectors, is difficult for cybersecurity technologies because organizations may be locked into a contract with a vendor and susceptible to the sunk cost fallacy. It is also important to avoid giving one vendor a monopoly or discouraging vendors from innovating by making it difficult to turn a profit. It may be that market forces and self-interest are sufficient in most cases as long as transparency into tool efficacy is encouraged and maintained. More research is needed to determine when and if forced recall of AI-based tools makes sense and how that process should be governed.

Limitations

Because the resilience of AI-based cyber defence tools depends on the robustness of the underlying algorithms and ensuring the robustness of such algorithms is an area of open research (Carlini, 2020; Goodfellow & Papernot, 2017), no vetting process can protect against every potential exploit because the space of potential exploits has not been enumerated. As noted by Carlini, "It's very difficult to differentiate between true robustness and apparent robustness, where true robustness means that no adversarial example actually exists, and apparent robustness means it looks that way because we haven't found one yet" (2020). However, that does not by any means suggest that vetting is useless.

Many of the cryptography algorithms used millions of times every single day are not known to be one hundred percent secure. There might be a vulnerability that has not yet been discovered or at least it is not publicly known. Furthermore, because these cryptography algorithms go through a rigorous and transparent vetting process, they effectively protect private transactions and data from most types of adversaries. Similarly, rigorous testing of AI-based cyber defence tools will go a long way in making our networks more secure, even if it does not ensure protection against every edge case or zero-day exploit.

Conclusion

The analogies, structures, and processes we present in this article are ultimately an exercise in formalization of quality assurance and testing practices inspired by another domain. In our case, that other domain is pharmacology and immunology. As with any other inspiration, it is simply that: an inspiration. We do not place any strong claim that the development and evaluation of new vaccines is an exact solution or a precise model for the problems of cyber defence reliability and resilience. However, the problems we highlight—namely, the critical need for the use of AI in cyber defence, and the inherent uncertainty about its ultimate resilience to adversarial exploitation—are real problems that need to be addressed. Furthermore, our early experiences as well as the rigor and discipline that we apply in this problem domain are valuable and worth formalizing. Our attempt to map our process to the process of drug discovery and vaccine safety is to propose a practice with even greater rigor. For that reason, we hope to inspire policymakers to observe some of the practices we propose and consider them as foundations for a future, formal process. The cyber threat and the risks of AI adversarial exploitation are real, and so is the need for a rigorous, transparent, disciplined, and staged evaluation of AI-based cyber defence tools.

Acknowledgments

We wish to thank reviewers that have helped polish this document. The research is based upon work supported by the US Department of Defense (DOD), Naval Information Warfare Systems Command (NAVWAR), via the US Department of Energy under contract DE-AC05-00OR22725. The views and conclusions contained herein are those of the authors and should not be interpreted as representing the official policies or endorsements, either expressed or implied, of the DOD, NAVWAR, or the US Government. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. ■

About the authors:

Edmon Begoli, PhD, is a distinguished member of the research staff with the Oak Ridge National Laboratory (ORNL) and is the Head of the Adversary Intelligence Systems Section. His research interests are in resilient and reliable system architectures for large-scale data analysis in mission critical domains. Edmon holds undergraduate, graduate, and doctoral degrees in Computer Science, and is an adjunct professor of Computer Science at the University of Tennessee-Knoxville, EECS department.

Kathryn Knight is a member of the Data Lifecycle and Scalable Workflows group within the National Center for Computational Sciences at Oak Ridge National Laboratory. She is also a Ph.D. candidate in the School of Information Sciences at the University of Tennessee. Her research interests revolve around information architecture and knowledge management, particularly related to how information is organized, represented, retrieved, and used.

Sean Oesch, PhD is a software architect turned security researcher. His recent research focuses on applications of ML to cyber and usable security. While at ORNL he has worked on a wide variety of projects, from cyber forensics to using ML to optimize fuel efficiency at stoplights. He completed his PhD at University of Tennessee, Knoxville.

Robert A. (Bobby) Bridges received a B.S. in 2005 from Creighton University and a Ph.D. in 2012 from Purdue University, both in Mathematics. Bobby joined Oak Ridge National Laboratory as a postdoc in 2012, was promoted to Associate Researcher in 2013 then Staff Researcher in 2017, and now serves as the Acting Cybersecurity Group Leader. Bobby currently leads a team providing ongoing scientific support to assist the US Navy in acquisition decisions through large-scale, scientific testing of market-leading cybersecurity tools.

References

ATT&CK evaluations. (n.d.). Retrieved February 09, 2021, from https://attackevals.mitre-engenuity.org/enterprise/evaluations.html?round=APT3

Bridges, R. A., Oesch, S., Verma, M. E., Iannacone, M. D., Huffer, K. M., Jewell, B., ... & Tall, A. M. (2020). Beyond the Hype: A Real-World Evaluation of the Impact and Cost of Machine Learning--Based Malware Detection. *arXiv preprint arXiv:2012.09214*.

Buchanan, B. (2020, August 10). *A national security research agenda for cybersecurity and artificial intelligence*. (Center for Security and Emerging Technology, May 2020), retrieved from cset.georgetown.edu/research/a-national-security-research-agenda-for-cybersecurity-and-artificial-intelligence/

Carlini, N. (2020). A (short) primer on adversarial robustness. Presentation at CVPR Workshop on Media Forensics, Seattle, WA.

Evans, S. R. (2010). Fundamentals of clinical trial design. *Journal of Experimental Stroke & Translational Medicine*, *3*(1), 19–27.

Jervis, R. (1978). Cooperation under the security dilemma. *World Politics 30*(2), 167–214.

Gartner for IT. (n.d.). https://www.gartner.com/en/information-technology

Glaser, C. L. (1997). The security dilemma revisited. *World Politics 50*(1), 171–201.

Goodfellow, I., and Papernot, N. (2017, February 15). Is attacking machine learning easier than defending it?" Cleverhans-Blog. Retrieved from cleverhans.io/security/privacy/ml/2017/02/15/why-attacking-machine-learning-is-easier-than-defending-it.html

Gore, A. (n.d.). How to think about security: asymmetry of difficulty. https://polyverse.com/blog/how-to-think-about-security-asymmetry-of-difficulty-498eeebe91b5/

Katzir, Z., and Elovici, Y. (2018). Quantifying the resilience of machine learning classifiers used for cyber security. *Expert Systems with Applications 92*, 419–429. https://doi.org/10.1016/j.eswa.2017.09.053

Levy, J. (2019). Open Source Vulnerability Databases. https://resources.whitesourcesoftware.com/blog-whitesource/open-source-vulnerability-databases

Li, G., et al. (2018). *Security matters: A survey on adversarial machine learning*. arXiv. https://arxiv.org/abs/1810.07339

McDermott, K. (2020, December 7). Winners of artificial intelligence challenge announced. Retrieved January 31, 2021, from https://www.navy.mil/Press-Office/News-Stories/Article/2436651/winners-of-artificial-intelligence-challenge-announced/

MacDonald, N., et al. (2020). Global vaccine action plan lessons learned I: Recommendations for the next decade." *Vaccine 38*(33), 5364–5371. https://doi.org/10.1016/j.vaccine.2020.05.003

The MITRE Corporation. (n.d.). https://attackevals.mitre-engenuity.org/

The MITRE Corporation. (2020). Researcher Reservation Guidelines. https://cve.mitre.org/cve/researcher_reservation_guidelines

Mulligan, D. K., and Bamberger, K. A. (2019). Procurement as policy: Administrative process for machine learning. *Berkeley Technology Law Journal 34*, 781–858. http://dx.doi.org/10.2139/ssrn.3464203

Myneni, S., Chowdhary, A., Sabur, A., Sengupta, S., Agrawal, G., Huang, D., & Kang, M. (2020, August). Dapt 2020-constructing a benchmark dataset for advanced persistent threats. In *International Workshop on Deployable Machine Learning for Security Defense* (pp. 138-163). Springer, Cham.

Naval Information Warfare Systems Command (NAVWARSYS). US Department of Defense. AI ATAC 3 challenge: Efficiency & effectiveness afforded by security orchestration & automated response (SOAR) capabilities. Challenge website. https://www.challenge.gov/challenge/AI-ATAC-3-challenge/

NAVWARSYS. (n.d.-b). Artificial intelligence applications to autonomous cybersecurity (AI ATAC challenge. Challenge website. https://www.challenge.gov/challenge/artificial-intelligence-applications-to-autonomous-cybersecurity-challenge/

NAVWARSYS. (n.d.-c). US Department of Defense. Network detection of adversarial campaigns using artificial intelligence and machine learning. Challenge website. https://www.challenge.gov/challenge/network-detection-of-adversarial-campaigns/

Schneier, B. (2021, January 5). The SolarWinds hack is stunning. Here's what should be done. Schneier on Security. www.schneier.com/essays/archives/2021/01/the-solarwinds-hack-is-stunning-heres-what-should-be-done.html

Ring, M., Wunderlich, S., Scheuring, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. *Computers & Security*, *86*, 147-167.

Shah, N., Ho, G., Schweighauser, M., Ibrahim, M., Cidon, A., & Wagner, D. (2020, August). A Large-Scale Analysis of Attacker Activity in Compromised Enterprise Accounts. In *International Workshop on Deployable Machine Learning for Security Defense* (pp. 3-27). Springer, Cham.

Song, L., Shokri, R., and Mittal, P. (2019). Privacy risks of securing machine learning models against adversarial examples. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (pp. 241–257). ACM SIGSAC.

Tung, L. (2021). SolarWinds attack is not an outlier, but a moment of reckoning for security industry. ZDNet. Retrieved January 31, 2021, from https://www.zdnet.com/article/solarwinds-attack-is-not-an-outlier-but-a-moment-of-reckoning-for-security-industry-says-microsoft-exec/

Uesato, J., O'donoghue, B., Kohli, P., & Oord, A. (2018, July). Adversarial risk and the dangers of evaluating against weak attacks. In International Conference on Machine Learning (pp. 5025-5034). PMLR.

Veeramachaneni, K., et al. (2016). AI^ 2: Training a big data machine to defend. In *Proceedings of 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS)* (pp. 49–54). Institute of Electrical and Electronics Engineers. 10.1109/BigDataSecurity-HPSC-IDS.2016.79.

Wang et al. (2007). Measuring the overall security of network configurations using attack graphs. In *Proceedings of the IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, Berlin, Heidelberg, 2007.

Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1–18). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372833

ANALYSIS

About Integrity in the Telecommunications Ecosystem – a Nokia Perspective

Dr. OLAF SCHULZ
GOVERNMENT RELATIONS, HEAD OF EUROPE AND STRATEGIC PROGRAMS, NOKIA

JULIA JASIŃSKA
HEAD OF INTERNATIONAL RELATIONS & TRADE POLICY, NOKIA

Dr. SIMON RINAS
HEAD OF GOVERNMENT RELATIONS GERMANY, NOKIA

MATHILDE MAURY
GOVERNMENT RELATIONS EUROPE AND STRATEGIC PROGRAMS, NOKIA

ABSTRACT:

5G introduces new technologies connecting critical infrastructures and enabling digital economies and societies, making security an imperative. Users must trust that their technology is secure, and will not be exploited by malicious actors. Thus, digital trust is an integral part of 5G's success. Trust builds on a comprehensive picture of conduct, based on values, bound by integrity. This article outlines the role of corporate integrity in supplying 5G networks and discusses why engaging in ethical business practices is as critical as securely designing technology to fulfil the 5G promise.

Keywords: 5G, digitalization, security, trust, integrity

The 5G Promise

5G and the technologies it enables promise to have a significantly productive impact on our societies. It is expected to extensively transform global economies by as early as 2030, delivering $8 trillion in value around the world ("Nokia's 5G Readiness Report," 2020 October). 5G will connect everyone and everything to each other, digitally transforming even the most physical aspects of our lives. Every industrial segment, public service or critical infrastructure will be touched by the 5G revolution.

5G promises to be much more capable, flexible, but also more complex than earlier network generations, using a heterogeneous architecture comprised of multiple access as well as physical and virtual infrastructure technologies. Many traditional network elements of 4G are replaced in 5G by Virtual Network Functions and cloud architectures. It is estimated that billions of devices will be connected to 5G networks over the coming years – with many of these devices being low power sensors, wearables, and small industrial devices. 5G is expected to increase wireless capacity by 1000 times and will connect seven billion people and seven trillion IoT devices around the globe ("5G Infrastructure Public Private Partnership," n.d.).

Yet, the more indispensable 5G networks become, the bigger the prize for malicious actors to interfere with them for commercial, political, or other reasons.

Because of how crucial the networks will become for the national economy and for national security, in 5G, security is imperative! The nervous system of the society must be reliable and well protected against any disruption. This is a task for all stakeholders involved – such a comprehensive challenge cannot only be tackled technologically.

Because of how crucial the networks will become for the national economy and for national security, in 5G, security is imperative.

Some non-technical aspects have been widely discussed recently in terms of supplier trustworthiness. Current and potential future dependencies that might force a supplier to conduct or enable malicious actions by leveraging its capabilities, access to, and unique knowledge of the networks concerned, have been identified as some of the most relevant criteria when assessing such trustworthiness. Further, deeming a company trustworthy also builds on addressing their attitude and actual conduct, and therefore their integrity.

A track record of illicit activities, ignorance and/or inaction towards product misuse leading to blatant violations of human rights, or employees poor working conditions are all individual signs of a questionable attitude. Given such lack of integrity, reasonable doubt can be raised regarding the capacity and/or intention of such companies to operate in a transparent and trustworthy manner – especially when put under pressure. Therefore, scrutinizing the integrity of providers of critical technology becomes essential for long-term security considerations.

Technology & Processes

International standards and protocols require the parties that install, maintain and operate telecommunications network infrastructure and services to ensure that appropriate technical and procedural safeguards are in place to protect the networks against attacks by malicious actors attempting to sabotage or manipulate the functioning of the network or parts of it, steal or otherwise compromise the data, or hold companies to ransom.

Realizing the diversity of 5G use cases will make securing the network even more complex. Availability, confidentiality and integrity of all users, management and control functions need to evolve to cater to dynamic networks, multiple players involved in service delivery, and a wide variety of devices, users, and applications. This complexity leads to an expanded attack surface. Moreover, the huge number of connected devices in the Internet of Things (IoT) also means that the network may be exposed to extensively spread attacks.

The 3GPP-specified network functions are created with a high amount of scrutiny providing a robust security design. But the deployment and usage are not necessarily mandated and therefore not always enabled. The mere existence of appropriate security considerations in standards may not be sufficient to achieve security assurance required by the criticality of 5G applications. Exploiting flaws in the design, implementation or configuration of the network cannot easily be completely avoided, especially inside the software implementing the network functions and in the configuration of complex network elements. Additionally, all external interfaces of the network may become subject to attacks. Therefore, security-by-design development processes and secure operational processes are key elements to technically secure network operations.

The mere existence of appropriate security considerations in standards may not be sufficient to achieve security assurance required by the criticality of 5G applications.

Trust

Despite all these actions, there are certain risks that cannot be prevented solely by technical means. These are caused by “malicious insiders”. Insiders may be those who have access to network elements and the related control functions or have a deep knowledge about the technology deployed in the field and how it is operated – or both. It could be employees or consultants of network operators, tenants and partners of the network operator, technology vendors, or managed service providers or subcontractors.

Trustworthiness in this context means that there are no reasonable grounds to assume that the privileged position of an actor, e.g. a company, in terms of access to and knowledge about a telecommunications network, its components and its management, could be leveraged in the future to conduct or facilitate “malicious actions”, such as undue access to data or manipulation of network functions.

This is no unusual consideration.

Background checks and trustworthiness assessments are routine, for instance in the defence sector. It is a precautionary measure to avoid risks – including second agendas or otherwise malicious intentions. It also involves vetting whether there is an advanced risk that an actor is, or could become, dependent on a third party with malicious intentions.

Background checks or trustworthiness assessments for providers of telecommunications infrastructure have been widely discussed in the context of 5G security policies (“Secure 5G deployment in the EU: Implementing the EU toolbox - Communication from the Commission,” 2020 January).

Integrity

Even a flawless technology can be misused in the wrong hands. For this reason, three elements should be taken into consideration: the technical and procedural network integrity, external political and legal circumstances, and also, the integrity of the actors involved.

Integrity is a comprehensive picture of conduct, based on values.

Most companies have a code of conduct requiring good behaviour from their employees. Yet, the virtue of integrity is to live by one’s standards: implementing policies and processes that steer all employees towards appropriate conduct, diligently reacting to any misbehaviour and creating a culture of compliance. While flaws are inherent to human conduct, addressing individual or corporate misbehaviour and drawing conclusions, thus being aware of one’s behaviour’s consequences, makes a difference.

Integrity becomes visible when difficult trade-offs occur. Human rights risks may occur in supply chain management; technologies which have considerable societal benefits may also be misused to infringe on human rights. Digital technologies could be used for surveillance or to intercept communications, limit freedom of expression, block access to information, and reduce the exchange of ideas. Maintaining the capacity to scrutinize business opportunities from a human rights perspective might often be a challenge in price competitive environments. But it can also become a competitive edge if societies decide to value integrity.

Most companies have a code of conduct requiring good behaviour from their employees. Yet, the virtue of integrity is to live by one's standards.

Integrity means also to act according to one's values, especially when it is neither easy, nor expected. Upholding the same standards everywhere, and at any time, is pivotal. A company supporting or turning a blind eye to the suppression of societies or corruption in one part of the world loses its overall integrity regardless of how vocal it might be about human rights in another place.

The Nokia Example

At Nokia, we create the technology that helps the world act together, connecting people to opportunities to improve their lives through access to education, enhanced healthcare, information and public services, market opportunities and more. The capabilities offered by 5G, the Internet of Things, cloud computing, Artificial Intelligence, Machine Learning and analytics are immense. Connectivity has never been so important as now. Acknowledging the responsibilities we bear, we build our business on a foundation of trust. It covers our approach to ethics, compliance and anti-corruption, ensuring responsible sourcing, respect for human rights and inclusivity, data privacy and security. It is our corporate commitment to earn this trust every day, in all of our business activities and in every country where we operate.

It is our corporate commitment to earn this trust every day, in all of our business activities and in every country where we work.

Design for security – securing products through their entire lifecycle

Security and privacy are embedded into all Nokia products. Our "design for security" process ensures that security is designed in and managed throughout a product's entire lifecycle, supported by a rich set of technologies, tools and procedures. Nokia runs a complex supply chain of trusted suppliers, which is designed with sustainability and responsibility in mind. Its resilience relies on implemented business continuity plans, providing flexibility and reliability to minimize disruption to operations in a time of crisis, as well as multi-sourcing (i.e. availability of alternative sourcing), and implemented processes to manage critical components inventories.

Nokia is additionally ensuring product security through:

- A holistic and lifecycle approach to security. This is visible in our security scorecard, which includes KPIs for product security, services security, regulatory security, IT security, supply chain security, assurance and security culture.
- Mandatory Nokia Design for Security requirements that apply to all Nokia products and services.
- Our contributions to improving the standards. We play an active role in key standardization bodies that are shaping the latest in security standards and best practices, including GSMA SECAG which defined NESAS (security assurance scheme for networks), GSMA Fraud and Security group, 3GPP SA3 (defining security standards for 5G), ETSI and other.

Moreover, Nokia helps operators to improve the security of networks through a wide-ranging portfolio of end-to-end security products, solutions and services, such as security risk assessments, security solution integration and managed security operations. Our solutions address risks related to misconfiguration of networks, lack of access controls, and exploitation of the Internet of Things, handsets and smart devices with best-in-class solutions ("Cybersecurity in the age of 5G technology," 2020). Further, Nokia's NetGuard Adaptive Security Operations are the telco market's most comprehensive Security Orchestration, Analytics and Response (SOAR) solution. Providing end-to-end security, the suite integrates audit compliance, privileged access, threat intelligence, network-based malware detection, and certificate management.

Integrity & trust – our commitments

Nokia is committed to the Cybersecurity Tech Accord and our CEO signed the Digital Declaration initiated by the GSM Association (GSMA) in April 2018. We supported the Paris Call for Trust and Security in Cyberspace right from the start in 2018 and actively participate in international organizations that promote the principles of those three initiatives, e.g. WEF, Information Security Forum – probably the largest non-profit information security organization, GSMA and others.

Just as others, we have signed and joined these initiatives. However, these commitments are backed up by concrete actions and engagements.

At Nokia, being a trusted partner for critical networks, we create the technology that helps the world act together. We commit to doing so in a responsible way, being guided by an imperative of integrity built on our adherence to the highest ethical and environmental standards, a culture of compliance, and our positive track record with regard to security, privacy and transparency. It is reflected in how we operate and in the solutions we provide. We take responsibility for our actions, making serious and verifiable efforts to minimize potential negative impacts of the technology we create.

We make sure that our solutions meet the highest expectations regarding technical robustness, security and privacy. Further, we verify that our exports – particularly those involving a potential higher risk to individuals' rights and freedoms – meet not only legal requirements, but also adhere to higher, self-imposed standards. The technology we create must uphold fundamental rights and ethical principles. New technologies as well as new uses of existing technologies are helping to keep societies connected, but they can also facilitate new or more severe forms of misuse. To prevent this, Nokia has a rigorous Human Rights Due Diligence process, covering both sales and R&D phases, to pre-emptively screen the use cases of its technologies. As such, we will never knowingly allow our products or services to be misused. Given the constraints on our planet and the environment, integrous business practices must encompass efforts to combat climate change and ensure sustainability. We are working every day to increase the energy efficiency of our products, to reduce the emissions from our own operations and finally to create technology that can help other industries and organizations reduce their emissions, resource use and eco footprint[1]. To back up these engagements and echo our commitment to supporting the European Union's digital transformation and green transition, Nokia has joined the European Green Digital Coalition as a Founding Member in March 2021.

Integrity in Nokia's operations is underpinned by our Code of Conduct, which is clearly defined, enforced, outlined in our yearly People & Planet report and externally assessed. We are transparent about our business conduct, performance, or ownership, and we report those on a quarterly basis.

We also take action to guarantee the sustainability of our supply chain, for instance by providing online workshops for suppliers and a series of awareness-raising webinars to counter modern slavery, conduct responsible sourcing, and promote health

[1] To know more on this, please see: https://www.nokia.com/about-us/sustainability/combatting-climate-change/

and safety amongst others. We evaluate our suppliers regularly to ensure that respective measures are implemented and effective.[2]

External assessment and recognition

Nokia exposes itself to external verification of our integrity posture by taking part in several sustainability-focused ratings, indices, and benchmarks on a continuous basis. In February 2021, we were named for the fourth consecutive year (2018-2021), and the fifth time overall, as one of the World's Most Ethical Companies by Ethisphere. Nokia's human rights approach was also audited by the Global Network Initiative. We were proud to be awarded the Human Rights Campaign (HRC) Corporate Equality Index score of 100 percent.

In December 2020, we were happy to see our work recognized as Top 10 in the inaugural World Benchmarking Alliance's Digital Inclusion Benchmark which measures how the world's 100 most influential technology companies are helping to advance a more inclusive digital society.

Finally, Nokia's selection in January 2021 as a technology provider and collaborator for the United-States' National Cybersecurity Center of Excellence (NCCoE) 5G Cybersecurity Project, which aims to provide enhanced cybersecurity capabilities built into the network equipment and end user devices, is an additional proof that both the industry and public agencies deem Nokia a trustworthy partner ("Nokia selected for U.S. Federal 5G Cybersecurity Project," 2021 January 14).

Trust and integrity: a responsibility

Given our outstanding position in the market, Nokia recognizes its responsibility as a company providing 5G networks around the world and as a market leader for telecom software – a crucial element given that virtualization and cloudification are the dominating trends in network architecture. Our customers include telecom operators and enterprises from energy, transportation, mining and many other sectors. As we help them on their digitalization journey, we understand our special obligations as a company providing these critical networks, with a deep knowledge and access to it. Our customers and the customers of our customers must trust that this is not exploited by any party with malicious intentions. Our integrity makes the difference. ■

For more information please see: Conducting our business with integrity | Nokia

[2] 332 supply chain audits were conducted in 2019, of which 45 were onsite audits on corporate responsibility topics.

About the authors:

Dr. Olaf Schulz is heading Government Relations Europe and Strategic Programs in Nokia. He ensures thought-leadership in critical areas like network resilience and leads strategic engagement with public and third-party stakeholders.

Julia Jasinska is heading International Relations and Trade Policy in the government relations team of Nokia. She is the leading trade policy expert and she supports Nokia's leadership engagements with governments, international organizations, and others on policy matters.

Dr. Simon Rinas is Head of Government Relations Germany at Nokia. He holds a PhD in political science and has held positions in telco companies, think tanks and universities. His focus is on EU and national digitalization policies, including 5G.

Mathilde Maury is a contractor for Nokia, appointed to the Government Relations Europe and Strategic Programs team. She focuses on global political developments and their strategic implications, on digitalization and network resilience policies.

References

5G Infrastructure Public Private Partnership (5G-PPP). (n.d.). https://5g-ppp.eu/about-us/

Cybersecurity in the age of 5G technology. (2020). https://onestore.nokia.com/asset/207438

Nokia's 5G Readiness Report. (2020, October). https://www.nokia.com/networks/5g/readiness-report/

Nokia selected for U.S. Federal 5G Cybersecurity Project. (2021, January 14). https://www.nokia.com/about-us/news/releases/2021/01/14/nokia-selected-for-us-federal-5g-cybersecurity-project/

Secure 5G deployment in the EU: Implementing the EU toolbox - Communication from the Commission. (2020, January). https://ec.europa.eu/digital-single-market/en/news/secure-5g-deployment-eu-implementing-eu-toolbox-communication-commission

ANALYSIS

The *Panoptes Maya*: Virtualizing the Foucaultian Panopticon in Surveillance and its Reinforcement Through the Adversarial Internet

ESTHER STAES
MA STUDENT IN INTERNATIONAL SECURITY AND DEVELOPMENT, JAGIELLONIAN UNIVERSITY IN KRAKÓW;
CHRISTOPHER A. WIJONO
MA STUDENT IN INTERNATIONAL SECURITY AND DEVELOPMENT, JAGIELLONIAN UNIVERSITY IN KRAKÓW

B
K
C
L
Q
I
G
D
A
H
E
10
20
30
40
50
60 Feet
N
M
F
O

CYBERSEC
YOUNG LEADERS

Acknowledgements to fellow MA students Marco Reyes Bernal and Ekrem Gökçe for their input and participation in the creation of this article.

ABSTRACT:

The Panopticon was developed by Jeremy Bentham as a system of prisoner control. Later philosophers like Michel Foucault would argue that the system goes beyond the prison and encompass greater society. We explore the implications of the Panopticon in the contemporary era and its creation of a disciplinary society through new technologies. While some functions of Foucault's Panopticon remain, there are noticeable changes to the nature of the panopticon itself and its influence on society, its control, and its surveillance.

Keywords: Panopticon, adversarial internet, (cyber) surveillance, disciplinary society

Opening Statement

The implications of technologically backed, overarching surveillance have been explored by multiple authors in the past. The suppression/elimination of individuality and technology's triumph was discussed by Zamyatin in *We* (1924) and Rand in *Anthem* (1938). Following these two works, technology's use as a tool of surveillance, control, and oppression was codified in the seminal dystopia that is George Orwell's *Nineteen Eighty-Four* (1949). Certainly, Orwell's vision was so striking and relevant to the age (like secret police, e.g. Stasi or NKVD) that "Orwellian" entered the cultural lexicon as a byword for totalitarian rule marked and enforced by total surveillance and control over action and thought, motivating studies on Orwell's brand of futurology and its parallels on contemporary society. This common thread of control over action and thought was explored prior by Bentham and his proposal of Panopticon (1791); through it, Bentham envisions a prison in which a single prison guard is capable of observing every inmate contained in the institution, without the inmates knowing whether or not they are being watched at any given moment, i.e. an "unequal gaze". In this research, however, we depart from Orwellian futurology and Bentham's architectural pursuit towards the philosophical exploration of the concept pioneered by French philosopher Michel Foucault. Here, we propose how the purpose and vision of panopticon has been or could be realised through adversarial internet, itself a product of technological advancements as discussed through the points of view of the aforementioned writers; how aspects of internet, technology, and connectivity (e.g. internet of things, backdoors, data collection, etc.) can be transformed, if not weaponised, towards realising this panopticon and its purported objectives. If we look at cybersecurity as "the body of technologies, processes, and practices designed to protect networks, devices, programs, and data from attack, damage, or unauthorized access" (De Groot, 2020), then a connection can be made; the concept of panopticon and its inherent position of hierarchy and power is suited for analysing the discourse of ethics/distinctions between legal and illegal, acceptable and unacceptable, belligerent or benign conduct (attacks, thefts, etc.).

Here, we propose how the purpose and vision of panopticon has been or could be realised through adversarial internet, itself a product of technological advancements as discussed through the points of view of the aforementioned writers; how aspects of internet, technology, and connectivity (...) can be transformed, if not weaponised, towards realising this panopticon and its purported objectives.

Foucault: Panopticon and the Disciplinary Society Virtualised

Foucault elaborated his position in *Discipline and Punish* (1975) and *Security, Territory, Population* (2008; originally lectures given in 1978). In the former, Foucault argued that the prison model of torture, punishment, discipline, and prison had encroached upon broader and broader aspects of society (forming disciplinary society), and in the latter, he explored the means by which the sovereign imposes and exercises this capacity/authority to discipline.

Therefore Foucault expands upon Bentham's panopticon; he posits that the disciplinary institution of the prison, the imposition of "unequal gaze", carries bigger implications than the simple observation of its subjects or the question of architecture. It allows for a (gradual) transformation of their thought and behaviour, through becoming one of the means the sovereign has to do so. In *Discipline and Punish*, Foucault explored the concept of Panopticon as a means to institute "discipline" and its application beyond prison; he argued that, for example, the military, government institutions, hospitals/sanatoriums, schools, churches, etc. are subject to this process. This leads to the creation of what he defined as "disciplinary society", a "society where one becomes a docile body due to the presence of constant surveillance". Docility here takes many forms: civil disengagement, conformity, or even, to borrow from Orwell, doublethink.

Therefore, focus must be placed on this "unequal gaze". Foucault proposes that there is an inherent hierarchy of the observer and observed where the observer maintains the capacity to survey, and this should and would have been extended to the capacity to transform and influence subjects as a result. Furthermore, in *Security, Territory, Population*, Foucault would argue that the observer is in and of itself empowered and authorised (i.e. governmentality) to conduct this surveillance upon its subjects, and to also impose punishment in the case of violations/deviations. On the question of power, the sovereign also maintains the capacity to dictate what is wrong or right, and what is allowed or otherwise; as for lawfulness, the state legitimises its monopoly on specific forms of surveillance and control. Here, Foucault draws upon the concept of national interest (French: *raison d'état*) and extends the be-all and end-all of the state towards surveillance.

On the question of power, the sovereign also maintains the capacity to dictate what is wrong or right, and what is allowed or otherwise; as for lawfulness, the state legitimises its monopoly on specific forms of surveillance and control.

Therefore, multiple research questions can be derived from the above, the encroaching of the Panopticon beyond real institutions and the virtual:

- How is the panopticon realised through contemporary technology, and adversarial internet?
- How does technology-backed panopticon influence the thoughts and behaviours of other people? Are there any discernible changes between the past and present?
- How does the sovereign/government justify virtual surveillance? Have there been any shifts in the power dynamics of observer and observed?

System Studies: Virtualizing the Panopticon in Adversarial Internet

Multiple research questions were proposed for deepening our understanding of both the panopticon and cyber surveillance. A system study of Bentham and Foucault's panopticon revisited in the contemporary era would be the answer to the first question: "How is the panopticon realised through contemporary technology, and adversarial internet?" To do so, we need to examine the similarities and differences between the panopticon and the contemporary society, with placing special focus on the notion of adversarial internet.

As stated, the panopticon's architectural legacy stands on two main features: the watchman and observed occupants. Presently, with the expansion of the digital realm, social media, and the unequalled amount of data circulating in the internet, decision-makers must deal with the tenuous dichotomy between security and freedom. Moreover, in the age where state apparatuses such as the National Security Agency and Government Communication Headquarters are given rein to maintain substantial capacity to survey and observe, the question of whether the panopticon is a useful way of analysing our societies remains. Foucault proposed in the '70s that the panopticon mirrors his idea of surveillance and enforcement in a disciplinary society. He argued that its construction in itself was "the culmination of technologies of power rather than their beginning" where a prisoner, recognising he could be watched at all times, begins to watch himself and becomes "the principle of his own subjection". This manner of "soft" power – self-induced power – turns a disciplinary blockade into a disciplinary device (French: dispositif), "from exclusion and blockade towards generalized discipline" (Murakami, 2007). Foucault also wondered why disciplinary progresses weren't as much celebrated. As Roy Boyne (2000) remarked on the trend:

> Any deep critique of surveillance as a principle would have to imply a critique of social democracy and social welfare simultaneously, and may help explain the relative calm with which the contemporary development of surveillance powers has been received.

He argues that presently, "social works" such as medicine, education, psychology, etc. assume an increasing share of supervision and assessments. Therefore, criticising surveillance would mean criticism of Western social democracy. An important assertion is that not everything is becoming a prison, but that the prison itself is losing part of its purpose.

As a result, the watchtower constituting one part of the panopticon is presently equivalent to the security cameras that aim to secure buildings and institutions; visible machines monitoring and watching events through human-like eyes without the human themselves present. For example, a citizen knows that he's being watched by security cameras in a supermarket. Hence the shop is a panopticon where the citizen psychologically watches his own actions, becoming "the principle of his own subjection" as much as he is subject to the proprietors and his observers. Do we see the same behavioural changes in the context of digital surveillance and data usage? In this case, whether it is via smartphones, social media accounts, geographical locations, etc., the user is not aware that they are being watched. In the post–Edward Snowden era, we are more aware of the use of our private data by not only different government agencies, but also private actors such as Facebook, Google, or companies like the former Cambridge Analytica. This data flow via adversarial internet (and non-adversarial means) is more and more acknowledged by citizens. However, because citizens aren't confined in cells (i.e. established physical boundaries), and there are no physical markers of their observation (the watchtower or security camera), they do not maintain this feeling of being watched. Most of the data collected is surrendered as soon as the user clicks the "I accept these privacy policies" button, even without proper reading of the legal statement. Would it be enough to prove that digital surveillance is a panopticon? It is important to recall that the primordial goal of the panopticon was to correct behaviour. Therefore, without the existence or illusion of this "unequal gaze", there would be no influence over thought and behaviour; what does this imply? Does it mean that we have been, or are increasingly inured to normalised in the way the panopticon was constructed? One of the answers to that question is that the surveillance role has moved from changing behaviour to bringing "security".

Most of the data collected is surrendered as soon as the user clicks the "I accept these privacy policies" button, even without proper reading of the legal statement. Would it be enough to prove that digital surveillance is a panopticon?

Furthermore, another difference is addressed by Jake Goldstein in McMullan (2017): the intangibility of data surveillance, and the detachment of data from the self, i.e. separation from our bodies/corporeal form. This is a departure from the panopticon, whose foundation lies on its employment of a physical construct/boundary (i.e. the eponymous prison) to govern physical conduct and maintain control over the human body. This is compounded by the fact that most users do not know the extent of the data they distribute and its collection, whether to private parties/corporations or governments. What is observed? The inability to answer that question makes it impossible for citizens to regulate their behaviour. But it remains that the role of the "digital" panopticon changed. The future projects an increase in automation and interconnection; the advent of the internet of things (IoT). Common everyday items such as refrigerators, air conditioning systems, automobiles, etc. will impact the future of surveillance because not only will they communicate with each other, but those transmitted data will find a way to corporations or governments, mostly through the use of adversarial internet. Hence, citizens will become even more scrutinised in the form of "asymmetrical digital panopticon" created to bring security.

Identifying Changes

While changes can be attributed to elements beyond that of communications technology and the adversarial internet, ICT nevertheless maintains a substantial role in making these changes possible.

Deleuze: Control Society

The above findings correspond to and necessitate revisiting the discussion on the abolishment of physical boundaries/institutions proposed by Gilles Deleuze in Society of Control (1992). Deleuze identified three main changes/departures from Foucault's disciplinary society. The first was the diminishing physical boundaries/institutions and the increasing prominence of intangible ones, some with overarching influence on multiple aspects of society. Here, Deleuze would already envision the computer as the central controller/manager of this society of control. A pertinent example pointed out by Deleuze is the financial system. For example, mechanisms such as moneylending/banking (debts), the use of fiat currency (itself issued by the state), and the market are all controllable and observable through computers and communications technology. The second is the transition from traditionally state-run institutions (predominantly the school and the military – tools the sovereign uses to institute the disciplinary society) towards increasing privatisation. Deleuze returns to the financial world and points out banks and corporations as an example. The third, and perhaps currently most relevant, is the transition from sanction (the use of penalties, etc. to dissuade deviant behaviour; practices pointed out by Foucault in Discipline and Punish as, for example, torture or punishment) towards reward – promoting favourable behaviour through the promise of positive gain. In the financial system, for example, banks offer favourable interest rates to get people to entrust their money to them, or the gradual transition towards cashless society by offering deals and benefits to get people to enter the system and, in the process, entrust money (and data) to the corporations.

Therefore, the transition between the corporeal and the incorporeal, supported by communications technology, has allowed the panopticon to transcend the illusory "unequal gaze"; rather than the single guardsman (traditionally the state) only being capable of observing some of the prisoners, the guardsmen (whose status is now increasingly blurred) are now, in fact, capable of observing, and controlling, all of them perpetually. The "prisoners" too are no longer bound in their physical holdings or separated by walls, and now enjoy some level of uncontrolled or unregulated communication. Furthermore, this also carries the implication that the intangibility means a blurred boundary between the isolated prison-panopticon and outside it; the panopticon's subjects may very well exist beyond the prison it purportedly controls, and at the same time we can assume that the panopticon's power to observe goes beyond its controlled prison as well.

The transition between the corporeal and the incorporeal, supported by communications technology, has allowed the panopticon to transcend the illusory "unequal gaze"; rather than the single guardsman (traditionally the state) only being capable of observing some of the prisoners, the guardsmen (whose status is now increasingly blurred) are now, in fact, capable of observing, and controlling, all of them perpetually.

And yet, as discussed, the idea of the very existence of this panopticon does not necessarily influence the observed in thought and behaviour. Can this perhaps be explained by the atmosphere of apathy, comfort, or benefit that the transition from discipline and punish to reward brings? Or perhaps, drawing from de Jouvenel's *On Power, the Natural History of Its Growth* (1948), they realise the possibility of placing themselves as "guards", the observers, and the capacity to influence that it brings? If anything, the transparency now means that citizens, the observed, are capable of influencing the means, mechanisms, and the legitimacy of the state's national interest (French: *raison d'état*) of surveillance. They are now empowered, or at least aware enough of their position, to argue for what's right and wrong, what's acceptable or not, and which lines governments and companies should not cross; what is adversarial and what isn't.

Proposed Framework: Atmosphere of Surveillance

Identifying these relevant elements: the agent, the method, and the reception allows a graph illustrating possible atmospheres of surveillance to be proposed. The graph will consist of a heptagon including 7 elements: Sentiment, Method, Agencies of Government and Private entities, Intensity of Sanction and Reward, and Domain. Each element is assigned a value of 0 to 1.

Sentiment represents the receptiveness of the population towards surveillance, ranging from negative to positive. A score closer to 0 represents greater negative reception, while a score closer to 1 represents greater positive reception; 0 represents total dislike for and opposition to the very idea of surveillance, while 1 represents total support towards surveillance, if not extended to hostility towards those against it. Middling scores would indicate multiple perspectives, such as regarding surveillance as a necessary evil. It can be discerned through multiple means, such as the evaluation of approval/trust ratings of institutions (government, corporations, etc.) and the people's circumvention of surveillance measures (the use of VPN, anonymity, etc.).

Methods represent agent accountability and the opacity/transparency of their conduct; opacity implies limited accountability through limited or no disclosure/justification of employed surveillance methods, while transparency implies open accountability and proper disclosure, justification, and explanation of surveillance and its rationale. A score closer to 0 represents greater tendencies of opacity, while a score closer to 1 represents greater tendencies of transparency. This can be evaluated through measures employed to communicate justifications/explanations, laws and regulations governing the process of surveillance, and the level of policing available to check and balance these measures.

Government/Private agency represents the extent of government/private entities' involvement and agency in multiple arms of surveillance, both the influence they have and the realms they are involved in. A score closer to 0 represents either minimal or partial involvement, while a score closer to 1 represents greater involvement. 0 would indicate no involvement at all and 1 would indicate overarching involvement on all aspects of surveillance. This can be evaluated through the engagement of their arms, both physical and virtual, the employment of technology to aid it, and the level of funding they receive.

Intensity of sanction represents the extent to which the government uses surveillance as a means to identify, deter, and punish bad behaviour. A score closer to 0 represents less interest/motivation in using surveillance in such a manner, while a score closer to 1 represents greater interest/motivation. It can be evaluated through available deterrents for bad behaviour, the percentage of recidivism, or the establishment of boundaries and grouping to prevent unwanted occurrences.

Intensity of reward represents the extent to which the government uses surveillance as a means to motivate, reinforce, and influence citizens to adhere to good behaviour and reward them accordingly. A score closer to 0 represents less interest/motivation in using surveillance in such a manner, while a score closer to 1 represents greater interest. It can be evaluated through the means citizens are conditioned to adhere to and the rewards the citizens receive; their nature and how it benefits them.

Domain represents the geographical reach the government has in exercising their capacity to survey and observe. A score closer to 0 represents smaller geographical boundaries, while a grade closer to 1 represents greater geographical boundaries. A score of 0 represents limited boundaries even in one's own state, and a score of 1 represents exercise at the global level. It can be evaluated through the reach and influence government instruments have in specific geographical areas.

The virtue of this framework lies in its capacity to be employed in comparative studies; for example, different governments may share equal values in some elements while maintaining different values in some. Adversarial internet can be used as a concomitant variation in which to evaluate these differences. Following consolidation of the elements and its evaluation, this framework would hopefully be of some significance in future quantitative and mixed methods research as an addition to prior qualitative research.

Towards the End of Presumption of Innocence?

The principle of innocent until proven guilty is a major legal regulation that was instituted in order to fight wrongful criminalisation. In the digital era and the increasing use of technology, it was thought that surveillance would aid its implementation. As Lord Steyn in Hadjimatheou (2016) argues:

> It is of paramount importance that law enforcement agencies should take full advantage of the available techniques of modern technology and forensic science. (...) It enables the guilty to be detected and the innocent to be rapidly eliminated from enquiries.

Therefore, the statement implies that the usage of modern technologies has become a necessity for the protection against wrongful suspicion or conviction of crimes.

Would that statement still stand when one takes adversarial internet into account? As a legal principle, the goal of presumption of innocence is to prevent the conviction of innocents. The purpose of this section is not to try to undermine the legal goal of this principle by claiming that surveillance results in an increase of miscarriages of justice per se, but on another implication: that presumption of innocence doesn't only result in a decrease of wrongful convictions, but also suspicions. If we analyse the principle through the moral assessment of innocence and trust, we can without a doubt assess that cyber surveillance and the usage of adversarial internet techniques cancels out the moral and philosophical definition of "presumption of innocence".

Hadjimatheou (2016) in her paper argues that there are no collections of data on individuals in liberal democracies that are as intrusive, communicative, or comprehensive as surveillance in prison. She goes further by stating that there are no corporations or individuals who have access to the entirety of the data collected on citizens.

Although it isn't false that we cannot identify one single entity that has *all* of the available data on individuals, it remains a half-truth; governments or corporations may not have a common data repository, but the ones available still represent a massive collection of intimate information on individuals. For instance, statements from journalists – such as Charlotte Guillard working for *France Télévisions* (2019) – accessing the Google database concerning themselves will probably not agree with the assumption that panoptic surveillance is only reserved for already proven-guilty criminals.

Furthermore, adversarial internet and its usage by governments and corporations remains an intrusive use of cyber surveillance. Willingly or not, it will have one major consequence: people will feel they're being watched and hence, being considered untrustworthy – even if the government's goal of security and discipline does not question the trustworthiness of their citizens. Therefore, we arrive at the claim that surveillance isn't waiting for someone to be allegedly involved in a crime to be observed. And since the goal of using surveillance for legal purposes is for "the innocent to be rapidly eliminated from enquiries", it should stand for conviction and suspicion; presumption of innocence in terms of suspicion cannot be valid while citizens are not aware of which data governments or corporations hold on them. And even if there are measures to ensure transparency in this process, the employment of adversarial internet implies that there remains some level of opacity; accordingly, the data-collecting entities risk distrust, anger, or opposition from the observed.

Furthermore, adversarial internet and its usage by governments and corporations remains an intrusive use of cyber surveillance.

Finally, the change from presumption of innocence toward a presumption of guilt could be explained by the fact that governments are still using surveillance to punish rather than to reward. That is if, as we explored, we had not observed a significant transition from sanction and punishment towards reward and control in the first place.

Future Projection and Recommendations

Following the discussion on panopticon, society and the adversarial internet, one question remains: "What could the European Union do as a supranational entity towards combating adversarial internet?" Answering the question requires prior analysis of two perspectives: the Chinese and the United States, epicentres of the biggest data collection and data communication in the world. On one hand, China is an authoritarian regime that imposed its cyber surveillance – and adversarial internet – use in order to control its citizens at the extreme. The Social Credit System, the Great Firewall, etc. create a country based on disciplinary society through Orwellian systems of surveillance. Moreover, that society is induced by the government but extends its reach to individuals through citizens as well. On the other hand, the environment in the United States is different in that it is a democratic country, meaning that the citizens are more empowered than in other regimes. It results in a cyber surveillance that is governmental yet maintains individual control: for example, the operations of the National Security Agency contrasted with the will of citizens to self-regulate a normalisation of the society; for example, the Black Lives Matter movement could be considered as self-disciplinary with its actions towards the censure of some cultural media, suppression of words, or even a sort of thought control. As much as the American citizens and government denounce the Chinese surveillance model as overwhelmingly collective and repressive, they themselves fall into the same Orwellian trappings of thought and behaviour control – the consequences of adversarial usage of the internet.

The European Union has already progressive privacy protection policies that are put into place: the GDPR (General Data Protection Regulation) laws allow the users to have a deeper control on which data corporations and governments could collect or not. Moreover, those policies allow the citizens to download the already collected data from corporations. However, three issues remain; those laws do not really prevent users from having their future data collected

– especially the data that will circulate through home devices in the future – corporations and governments won't release the data collected through adversarial techniques, and most importantly the data are still the exchange currency in the online contract between the user and the service provider either by private entities or the government. The citizens aren't aware of that because they do not reimburse the service through the exchange of traditional currency (i.e. money). Hence, while the used currency is data, there is currently no other way for corporations to profit or otherwise sustain operations without collecting private information on their users. One way to get out of that vicious circle would be to create partially paid internet services with traditional currency. Although an audacious idea that would probably not gain popular vote, it would allow citizens to regain a measure of control over their data and its collection while simultaneously sustaining services afforded by private entities.

Hence, while the used currency is data, there is currently no other way for corporations to profit or otherwise sustain operations without collecting private information on their users.

This solution would perhaps only prevent or otherwise mitigate normal data collection and wouldn't prevent the consequences of adversarial internet. In order to do so, the solution would be for governments to refrain from its usage. The more a government is using adversarial means to surveil its citizens, the more the users are trying to circumvent the regulations and surf the web anonymously, for instance using VPNs (Virtual Private Networks). The consequence would be that actual criminals could find a way to avoid observation and hence render surveillance obsolete. Eliminating the need for backdoors and the implementation of proper surveillance would allow it to once again regain its original purpose; to convict the guilty and exonerate the innocent. However, that would only be possible if governments reduce the usage of adversarial internet measures to a minimum; something the European Union should aspire to.

Concluding Remarks

The panopticon has changed, and with it, society. While the status of the guardsman has become increasingly blurred into a chimeric combination of government-private arms, the departure from physical boundaries and the observed' s increasing knowledge of their status within the panopticon (i.e. transparency) has allowed greater capacity for them to take increasing agency within it and stake a claim. And it is exactly the knowledge of this position that allows us to recognise the abuses and dangers inherent to adversarial internet; and to stand together against it. ■

About the authors:

Esther Staes is a Belgian student currently studying at the International Security and Development Master program of the Jagellonian University in Krakow. She received her bachelor's degree in Political Science from the University of Namur in Belgium and her research interests lay in political theories, corporation crimes, international relations, security and the psychological and philosophical insights for all of those interests. She helped create the Student Association of International Security and Development and is currently serving as its General Coordinator.

Christopher A. Wijono is an Indonesian student currently pursuing an MA in International Security and Development at the Jagiellonian University in Kraków, Poland after receiving a bachelor's degree in social sciences from the Ritsumeikan Asia Pacific university in Beppu, Japan. His research interests include political and international relations theory, political philosophy, media studies, and discourse on popular culture. He is currently serving as an inaugural leading member of the Student Association of International Security and Development, responsible for the Academic Affairs arm of the organization.

References

Bentham, J. (1791). *Panopticon or the inspection house* (Vol. 2).

Boyne, R. (2000). Post-Panopticism. *Economy and Society*, *29*(2), 285–307.

Blevis, F. and Guillard, C. (2019). *Ok Google… dis moi quelles sont les données que tu caches sur moi…* [News report]. France 3.

De Groot, J. (2020, June 10). What is Cyber Security? Definition, Best Practices & More. Retrieved July 14, 2020, from https://digitalguardian.com/blog/what-cyber-security

De Jouvenel, B. (1949). *On Power: Its Nature and the History of Its Growth*.

Deleuze, G. (1992). Postscript on the Societies of Control. *October*, *59*, 3–7.

Foucault, M. (1975). *Surveiller et punir.* Paris: Gallimard.

Foucault, M. (2007). *Security, territory, population: Lectures at the Collège de France, 1977–78*. Springer.

Hadjimatheou, K. (2016). Surveillance Technologies, Wrongful Criminalisation, and the Presumption of Innocence. *Philos. Technol.* 30, 39–54 (2017). https://doi.org/10.1007/s13347-016-0218-2

McMullan, T. (2015, July 23). What does the panopticon mean in the age of digital surveillance? Retrieved July 14, 2020, from https://www.theguardian.com/technology/2015/jul/23/panopticon-digital-surveillance-jeremy-bentham

Murakami Wood, D. (2007),. Chapter 23: Beyond the Panopticon? Foucault and Surveillance Studies. In S. Elden, J. W. Crampton (Eds), *Space, Knowledge and Power: Foucault and Geography*, pp. 245–264.

Orwell, G. (1941). *Nineteen Eighty-Four*. Secker & Warburg.

Rand, A. (1938). *Anthem*. London: Cassell.

Zamyatin, Y. (1993). *We*. London: Penguin Books.

ANALYSIS

In Cybersecurity, We Need Education and a Systemic Approach

ANDRZEJ DOPIERAŁA

PRESIDENT OF ASSECO DATA SYSTEMS, VICE PRESIDENT OF ASSECO POLAND

ABSTRACT:

The COVID-19 pandemic has changed the way companies and state administrations operate because they have had to adapt quickly to the new reality. In order to survive and have a chance for further development they have moved to the digital reality. We have observed not only the changes related to the transition to remote work, but above all a swift transformation of business processes in logistics, sales, or customer service. We have also had to look again at cyber security, which has been growing in importance for a long time and should become a priority in these times.

Keywords: cyber-security, cybercrime, ransomware, secure software

The notions of digital transformation or digitisation of the economy have been in the headlines recently. The COVID-19 pandemic has changed the way companies and state administrations operate because they have had to adapt quickly to the new reality. In order to survive and have a chance for further development they have moved to the digital reality. Some had already used digital solutions before, others have been forced to implement them in an accelerated mode. We have observed not only the changes related to the transition to remote work, but above all a swift transformation of business processes in logistics, sales, or customer service. The most important challenge has been to maintain business continuity. Along with digitisation, we have also had to look again at cybersecurity, which has been growing in importance for a long time and should become a priority in times of digital acceleration. Already half of humanity operates in digital space. When so many people do so many things in cyberspace, there will always be those who want to take advantage of others. The weak link in the digital world are most often human beings – on the one hand, they are an object of attack by cybercriminals and, on the other hand, the attack itself most often happens by their mistake.

Not Only COVID as a Change Catalyst

Coronavirus has reminded everybody of cybersecurity, for which, up until recently, many companies lacked time, sometimes budgets, imagination, or simply patience for this subject. According to the report "Cybersecurity Trends 2020" prepared by Xopero in 2019, i.e. even before the pandemic, one organisation out of three experienced a data breach. Ransomware attacks, cyberspying, or data leaks are no longer rare events, but a common occurrence. Websites monitoring security incidents record new cases of attacks every day. Even technology giants, which devote huge resources to cybersecurity, are not able to protect themselves from them. The pandemic has caused an increase in the number of fraud incidents in the countries most affected by coronavirus. Where there is chaos and anxiety, criminals have even more room for their actions. A clear surge in the number of phishing attacks was recorded in Italy, among others. In April, the World Health Organization recorded a five-fold increase in cyberattacks. The British Financial Conduct Authority published data showing that the number of financial fraud activities was decreasing, but this drop does not apply to fraudulent activities on the Internet, in which case the trend is the opposite. In turn, Google, which in April 2020 blocked around 258 million COVID-19-related emails daily, claimed that the number of phishing messages during the pandemic did not increase, but their topic changed. 18 million emails contained malware or attempted phishing, and 240 million were simply spam. In addition, IoT, artificial intelligence, machine learning have opened the door to a whole new world not only for their users but also for criminals. These phenomena have changed our approach to cybersecurity.

According to the latest PMR report on the cybersecurity market in Poland, the total expenditure on software, hardware, and services in this area oscillates around PLN 1.5 billion. In the next 5 years, another PLN 1 billion is to be won by suppliers. Companies usually indicate cybersecurity as one of their priorities, but this is not always followed by concrete steps – specific expenses and separate budgets. The research also shows that 66 percent of large companies in Poland have not implemented or have only an informal program of information exchange about threats, nor do most companies have a SOC team to monitor potential cyberattacks. However, increasingly more often the target of the criminals are not multinational companies, but small and mid-sized businesses, which are also more willing to pay the ransom. The most frequently attacked sectors are still as follows: health, finance, energy, manufacturing, and new technologies – these companies regularly become victims of ransomware and data leaks.

> **The pandemic has caused an increase in the number of fraud incidents in the countries most affected by coronavirus. Where there is chaos and anxiety, criminals have even more room for their actions.**

Organised Cybercrime Groups

Up until recently, hackers have been associated with self-taught, talented, young people, often loners who, in the privacy of their homes, break into other people's computers and perform operations that lead to more or less measurable losses to the victim. The concept itself was created in the 1970s or, according to some sources, even in the 1960s. In 1980, an article on people's computer addiction published in *Psychology Today* used the term "hacker" in the title: "The Hacker Papers".

Unfortunately, from innocent fun, hacking has turned into illegal cybercrime and a business worth billions of dollars. It is backed by people who, for different reasons, financial or ideological, or in order to collect information for various purposes, hack into individual computer systems and even take over entire ICT networks. Hackers not only create malware, but also use psychological tricks to get users to give out personal information, click on a malicious link or open an e-mail attachment containing such software. In addition to social engineering and malvertising, hackers often employ botnets as an infrastructure to conduct their business, take control of browsers as a vector of attack, and use Denial of Service attacks to paralyse the victim's online services. Malicious software is an element of almost every attack and in this respect, ransomware has been the infamous leader in the last several months – blackmail related to encrypting valuable data has unfortunately become a very effective way for cybercriminals to illegally enrich themselves.

In addition to typical financially motivated cybercrime, the so-called "state-sponsored" activities have grown in popularity in recent years and have now become the most dangerous form of attacks. They are inspired and, increasingly more often, also directly carried out by state organisations. Their perpetrators act on behalf of a particular state and carry out economic thefts or intelligence operations to destabilise another state. They are very determined and have enormous resources to conduct their activities. While mentioning political motives, it is also worth recalling hacktivism when the perpetrators act for political, social, or ideological reasons. And here we can give the examples of the attacks by Anonymous or LulzSec.

Still, the most common actions of hackers are attacks on multinational companies, which now happen so often that we will soon approach them as if they were classic crimes, taking place every day, and we will hardly pay any attention to them. An example from summer 2020 is the attack on the American manufacturer of smart watches and fitness bands Garmin. For several days, no Garmin service, including a physical activity monitoring application, ran for users around the world. The company's website and customer service were also unavailable. Media also reported that production lines in watchmaking facilities were blocked. Another example is the incident that happened to FireEye, the world's leading cybersecurity company based in Silicon Valley. For years, it has been an important partner for government agencies and global companies, not only in the United States, in combating the serious network attacks that these organisations have suffered. In last December, the company itself was the victim of an attack, which shows that there is no institution in the world that can feel 100% safe. In the same month, the victims of a cyber attack were iPhones belonging to journalists of the Qatar's Al-Jazeera television, although up until recently it has been said that the iOS system cannot be broken. In this case, specialists from the University of Toronto have drawn up a report which shows that malware was used to record conversations with the microphone in the hacked smartphone, take photos, and determine the position of the phone owner. Everything was happening without the owners' knowledge. So the examples can be multiplied. Only after this kind of incidents related to a serious security breach are organisations slightly more inclined to employ staff who are professionally capable of dealing with the topic of cybersecurity, with CISO (Chief Information Security Officer) at the forefront, which results in better strategic planning and more time being devoted to security issues at management board meetings.

Unfortunately, from innocent fun, hacking has turned into illegal cybercrime and a business worth billions of dollars. It is backed by people who, for different reasons, financial or ideological, or in order to collect information for various purposes, hack into individual computer systems and even take over entire ICT networks.

Systemic Changes and Security as a Team

Most companies and institutions are aware that some areas of their business and assets are related to or dependent on cybersecurity. However, this

knowledge does not always actually cover the areas most vulnerable to attacks. Low awareness of potential events and ignorance of threats are still common. The lack of a holistic view of the areas that may be targeted by a potential attack is an elementary mistake that prevents the design and implementation of an effective solution, and more advanced tools require full coordination and cooperation of structures responsible for cybersecurity, IT, and business. This is how cybersecurity is approached by few, usually the largest organisations.

There are many reasons for not taking sufficient care of network security. The most common are the lack of support from the management, low risk awareness, dispersed liability and the lack of a cybersecurity head in the organisation. Often the overall responsibility for this issue is shared and implemented by the IT department, which does not always have a full picture of the business conducted by the organisation and is not able to correctly define all priorities. In many cases investments are conducted chaotically and inefficiently. The company uses many different solutions at the same time and does not look at the problem as a whole - it does not plan any strategic solutions to prevent threats, but acts only reactively. The turning point in its approach to security is only the incident related to its violation. If someone is more fortunate, the situation ends only with fear and no consequences. However, in many cases an attack has a negative impact on business, involves losses that are both easy to estimate (loss of revenues, ransom costs, or financial penalties) and difficult to quantify, such as stopping business continuity, and in extreme cases even leads to the company's liquidation. It is positive if such an incident finally opens the eyes of managers to many things and makes them look for solutions.

It is therefore very important for companies and management to be aware that effective safeguards that could counteract incidents involve costs that are disproportionately lower than the losses incurred as a result of an incident that occurred. That is why it is so important to have a systemic and holistic approach to cybersecurity management and at the same time to choose the right solution provider and partner for this task, who should understand this systemic approach and think comprehensively about customer security. We cannot forget about teamwork, which is also an important guarantee of the success of the whole undertaking, both the cooperation within the organisation between individual departments and the agreement and mutual understanding between the company and the supplier. When this happens, ensuring security becomes easier and more effective, and the "security as a team" approach brings results.

If someone is more fortunate, the situation ends only with fear and no consequences. However, in many cases an attack has a negative impact on business, involves losses that are both easy to estimate (loss of revenues, ransom costs, or financial penalties) and difficult to quantify, such as stopping business continuity, and in extreme cases even leads to the company's liquidation.

Secure Software Is Key

One of the most important elements of the organisation's secure infrastructure is secure software. As a frequent target of indirect and direct attacks, the most important thing for an organization when choosing software is to trust the supplier, who should have experience, reputation on the market, and the ability to provide the appropriate quality of service. There are about 30 reputable global providers of cybersecurity solutions. On the local markets there are national manufacturers, offering tailor-made software and boutique services, usually for large customers with extensive systems. Such suppliers also implement ready-to-use programs from global manufacturers. The product should be selected according to the specific needs and expectations of a given customer.

Whatever solution we choose, it is important for the supplier to include the best and widest possible testing in the project. Particularly important is to conduct several tests by external parties. On the one hand, we have domestic suppliers offering solutions written specifically for a given client, well tested by several entities, developed and updated on an ongoing basis. With such software we also have the ability to catch functional errors, and in case of critical errors the update is carried out in express mode. For the needs of some industries and strategic customers, such products are sold together with the source code, so that the customer can track the system on their own, catch any loopholes in the system, and react accordingly. In turn, using solutions offered by global suppliers gives us the certainty that they have been tested by thousands if not millions of users. However, due to their widespread availability, anyone can buy them for a few dozen dollars, so they are potentially more vulnerable to attacks from organised groups that have easy access to them, search for vulnerabilities, and can attack entities that also use them. On the other hand, manufacturers care about security and are ready to pay big money, even to casual cybersecurity enthusiasts, when they point out a software loophole or error.

Schools Thrown into Deep Water

The transition from schools to remote teaching is a great example of digital transformation. We adapted to this new reality very quickly thanks to technology. The digitalisation of Polish schools has long been a necessity, the implementation of which has encountered many problems. The situation was changed only by coronavirus, which forced an express change in this area. The modernisation of educational institutions is a huge challenge for the government and local government administration, but also for the society itself – teachers, students, and parents. Until now, everyone has focused largely on infrastructure, fast Internet connection in each school and provision of computer equipment. With the challenge of remote learning, the need for even greater emphasis on the issue of cybersecurity on many levels has emerged.

Speaking of the system's cyber-resistance to remote education, it is also worthwhile to take advantage of this moment of change and take care of cybersecurity education as well. It should be introduced from the very beginning – from the moment children and young people come into contact with digital services. The issues related to the threats posed by the use of the Internet should be developed at the earliest possible stage of education, preferably to be included in the core curriculum of younger classes in elementary school, and complete basics maybe even at the preschool stage. Proactive rather than reactive measures should also be taken, data protection strategies should be developed, including user education, implementation of appropriate technologies and corrective actions. By training educational staff in threat protection, implementing the right infrastructure, and applying appropriate corrective action protocols, an educational institution can significantly increase the resilience of its IT environment to the most dangerous attacks and their consequences, such as ransomware, data loss, financial loss, and reputation damage. People are often the weak point in the entire security structure. It is much easier for cybercriminals to use psychological trickery and persuade an individual to click a dangerous URL to access the system than to break a firewall or other technological security measures. When many employees with access to sensitive data work remotely, often outside the security system of their company or organisation, this problem becomes extremely important.

Increasing the cyber-resistance in a company or institution is a continuous process. Cybercriminals are getting smarter and smarter, using more and more sophisticated methods to achieve their goals. Therefore, users must be ready for constant caution and continuous improvement of security methods. ■

About the author:

Andrzej Dopierała – President of the Asseco Data Systems SA Management Board, Vice President of the Asseco Poland SA Management Board.
Graduated from Warsaw University of Technology, Faculty of Electrical Engineering. He has spent several years in USA and Canada and after returning to Poland, from 1994 to 2006, he held various management positions at Polish branch of Hewlett-Packard. He became the HP Poland's President in 1998. In 2006 he left HP Oracle Corporation and accepted the role of President of Oracle Polska. While leading Polish entity of Oracle he was also holding regional responsibilities.
In September 2013 he joined Asseco Group assuming the VP position, responsible among others for Asseco Systems company, which in January 2016 was transformed to Asseco Data Systems SA, merging 6 companies from Asseco Group in Poland. He acts as the President of Asseco Data Systems, one of the biggest Polish IT companies and is responsible for electronic signature and trust services solutions. Furthermore, he serves as Vice President of Asseco Poland, responsible for Defense Sector and Cybersecurity Solutions. In April 2021 he assumed the role of Chairman of the Supervisory Board of ComCERT SA, Asseco Group company dedicated to cybersecurity related services.

ANALYSIS

EU's Artificial Intelligence Regulation: Facing the Challenge of a Technologically Neutral Policy

THEODOROS KARATHANASIS

PHD CANDIDATE IN EUROPEAN LAW,
UNIVERSITY OF GRENOBLE ALPES

CYBERSEC
YOUNG LEADERS

ABSTRACT:

Technologies with higher perceived risk and uncertainty may be also worse candidates for technology-neutral policies. Regulating AI systems is still only vaguely outlined, raising questions about future application of the principle of technological neutrality to such technologies. Making crucial the achievement of a stable and sustainable regulation of artificial intelligence, this article argues for a right balance between technological neutrality and market prosperity through a "technology-specific neutrality" approach.

Keywords: technological neutrality, Artificial Intelligence, EU regulation, legal responsibility

Introduction

The principle of technological neutrality requires that the law generate the same effects regardless of the technological environment in which these standards apply. Used as a core element of any regulation addressing technology (Gagliani, 2020), it may foster competition, transparency of public policies, and flexibility for industry innovation and evolution (The International Bank for Reconstruction and Development et al., 2011, pp. 203–204). However, the principle of technological neutrality may lack specificity and can produce undesirable consequences for its application in practice. As Rajab Ali points out (2009, p. 9), the "technological neutrality of the law can lead to regulations whose meaning is so vague that its application to technology is often a matter of conjecture".

> **The principle of technological neutrality requires that the law generate the same effects regardless of the technological environment in which these standards apply. Used as a core element of any regulation addressing technology, it may foster competition, transparency of public policies, and flexibility for industry innovation and evolution.**

In announcing an ambitious digital strategy for artificial intelligence, the European Commission has defined, in its White Paper on Artificial Intelligence (AI), the framework for Europe's digital future (European Commission, 2020b). The aim of the strategy is to make digital transformation beneficial for all, with a focus on people and society as a whole. The proposals presented aim thus to create a unique data market and to ensure the anthropocentric development of AI, as the first steps towards achieving these goals. The wide expansion of AI applications creates unprecedented opportunities and leads to the radical transformation of the economy, society, and the state, giving momentum to growth, but also giving rise to major changes in the labour market. But many issues remain unresolved, notably around the legal qualification and liability of AI systems or even the interpretation of the criteria for "high-risk AI".

The regulation of emerging technologies such as AI is partly controversial, due to the European Union (EU) effort toward a technologically neutral regulation. The following question arises therefore: Does the application of the principle of technological neutrality in European digital technology projects may effectively address future AI technology impacts?

The aim of this article is to highlight the need for a right balance between technological neutrality and market prosperity. If technological neutrality is a key principle of European regulation related to digital technologies; AI systems' legal liability encourages a necessary reorientation of the concept of technological neutrality. The article is divided into four sections. First, I present the poly-semantic interpretation of technological neutrality principle. While the concept of technological neutrality is posing a challenge to the European AI policy, its meaning is not immediately clear. Second, I analyse the principle of technological neutrality within the European AI policy as regards its upcoming legal regime. The AI systems' legal responsibility will serve as a case study in a third section. Finally, last section will provide an alternative choice as a policy prescription to the EU actors.

The Principle of Technological Neutrality, a Poly-Semantic Interpretation

In the field of electronic communications, the 2002 European Framework Directive made technological neutrality one of the guiding regulatory principles for the telecommunications sector in the EU (Directive 2002/21/EC). It is defined by its recital 18 as follows: "The requirement for Member States to ensure that national regulatory authorities take the utmost account of the desirability of making regulation technologically neutral, that is to say that it neither imposes nor discriminates in favour of the use of a particular type of technology, does not preclude the taking of proportionate steps to promote certain specific services where this is justified (...)." The principle of technological neutrality is widely accepted and seemingly clear, but, in fact, poorly understood. Therefore, it is important to clarify the principle's substance.

The concept of technological neutrality has emerged as a regulatory principle, where states are invited to take it in consideration. This concept implies that regulations can and should be developed in such a way as to be "independent of any particular technology, without promoting or discriminating against specific technologies as they emerge and evolve" (Craig, 2013). Neutrality and non-discrimination in the law are almost always "laudable objectives" (Craig, 2013). While regularly invoked as a regulatory starting point in policy documents around the world, they are, however, generally not sufficiently justified.

In 2006, Bert-Jaap Koops explained that, following the context, technological neutrality can have four main legislative objectives. It may have as its objective (a) the achievement of particular effects, in terms of the behaviour of peoples or the results of activities; (b) a functional value between different modes of activity, particularly offline and online; (c) a non-discrimination between technologies with equivalent effects; or (d) the sustainability of the law through efficient regulations (Koops, 2006, pp. 83–90). Other scholars distinguish three general meanings of the term.

According to Maxwell and Bourreau (2014, p. 1), technological neutrality may firstly mean "that technical standards designed to limit negative externalities (e.g. radio interference, pollution, safety) should describe the result to be achieved, but should leave companies free to adopt whatever technology is most appropriate to achieve the result". Secondly, technological neutrality may mean that "the same regulatory principles should apply regardless of the technology used. Regulations should not be drafted in technological silos". For example, the European Parliament and Council Directive 2000/46/EC of 18 September 2000 on the taking up, pursuit of and prudential supervision of the business of electronic money institutions aimed to be neutral in terms of the implementation between the various electronic money technologies, even if it was unable to achieve this objective.[1] Last but not least, there will also be circumstances in which regulators may use the threat of future regulation as an incentive to push the market towards self-regulation or co-regulation solutions, which may be more effective than repressive regulations. In such case, the principle of neutrality may also mean that "regulators should refrain from using regulations as a means to push the market toward a particular structure that the regulators consider optimal" (Maxwell and Bourreau, 2014, p. 1). But all these meanings or aims of technological neutrality may sometimes intertwine inextricably.

[1] For more detail, see Caresche (2012).

There will also be circumstances in which regulators may use the threat of future regulation as an incentive to push the market towards self-regulation or co-regulation solutions, which may be more effective than repressive regulations.

Thus, the adoption of technologically neutral provisions by the EU appears to be the way forward to address the unpredictability of technological developments and, therefore, to ensure that the law is sustainable to successfully respond to such unpredictable developments over a sufficiently long time period. At EU level, the concept of technological neutrality has found its place in many important initiatives related to new technologies, the field of AI among them.

Technological Neutrality and European AI Policy, Dealing with an Upcoming Legal Regime

The rules and principles of the GDPR, such as the concept of identifying the person concerned, are flexible enough to cover future technological developments and provide a sustainable protection. However, is the GDPR sufficient to deal with AI? May AI systems be controllable by and compliant with this regulation?

A recent case allows us to highlight how the application of a technologically neutral European norm to new technologies may be possible. A Dutch court recently ruled against an AI system using identification system called SyRI (System Risk Indicator), due to data privacy and human rights issues (The Hague District Court, 2020). The SyRI has been used by four cities in the Netherlands to identify people whose social benefit claims should be examined further. In particular, it collected information from 17 different government data sources, including tax records, marketing authorisations, and the land registry. Although the Hague District Court considered the use of new technologies to control fraud acceptable, it found that SyRI was too intrusive and violated the confidentiality guarantees granted by European human rights law as well as the "technologically neutral" 2016 General Data Protection Regulation or GDPR (Regulation (EU) 2016/679). As the rules and principles of the GDPR, such as the concept of identifying the person concerned, are flexible enough to cover future technological developments and provide a sustainable protection. Although the decision was made by a trial court and may be appealed, it is likely to set an important legal precedent within the Union for future applications of AI systems. Following European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, it is stated that "the Union must address the various aspects of AI by means of definitions that are technologically neutral and sufficiently flexible to encompass future technological developments as well as subsequent uses." The application of the principle of technological neutrality, understood at EU level as non-discrimination in favour of certain technologies, can also be an important constraint. According to the Commission's White Paper, the security and responsibility implications of AI include, for example, that people "having suffered harm caused with the involvement of AI systems need to enjoy the same level of protection as persons having suffered harm caused by other technologies" (European Commission, 2020b). However, the integration of software, including AI, into certain products and systems can change the way they work during their life cycle. In other words, consumer protection may differ depending on whether AI technology is used in products or services, whether it has been integrated into the product with its marketing or after, by the producer or by a third party.

AI systems cannot be defined clearly and unambiguously. In its 2019 report study on achieving a shared common knowledge of AI, the European Commission's high-level expert group defines AI systems as "software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal" (High-Level Expert Group on Artificial Intelligence, 2019, p. 6). This definition is thus so broad that it "would virtually include any system or process capable of collecting data, processing it, and acting based on this information" (Lozano & Murillo, 2020). Therefore, we must not ignore the risk that the vagueness which characterises certain terms and concepts could lead over the years to great divergences in the interpretation of the law and, consequently, to legal uncertainty. Technological neutrality must be therefore followed by significant specific regulatory constraints and the case of legal liability for AI systems invites us to rethink the concept.

The integration of software, including AI, into certain products and systems can change the way they work during their life cycle. In other words, consumer protection may differ depending on whether AI technology is used in products or services, whether it has been integrated into the product with its marketing or after, by the producer or by a third party.

AI Systems and Legal Responsibility – the Illustration of Risk in a Technologically Neutral Approach

The European Commission's publication of the White Paper on Artificial Intelligence and the European Parliament's resolution on automated decision-making processes provide an idea of what EU regulation of AI systems might look like in the long term. However, if the European Parliament's resolution is adopted as such by the EU, questions will remain unanswered about the legal responsibility of artificial intelligence. Indeed, legal responsibility for AI systems illustrates the complexity and risks of a technologically neutral approach.

Some of the generally associated problems with AI, such as the lack of transparency or unpredictability of concrete individual results, do not apply to all forms of AI systems, but rather to data-based AI systems where causality may be more difficult to identify. The unique feature of AI systems is that they not only operate on the database embedded in their software, but acquire, through data processing, "experience" that allows them to more effectively manage the tasks assigned to them. In this way, however, it becomes difficult to know to what extent the effects of AI systems on the outside world will depend on human influence (Karanasiou & Pinotsis, 2017). The question therefore arises whether the artificial intelligence system can be classified as a subject of law with rights and obligations, comparably to physical ones.

The Commission's working paper of the 27 November 2019 on *Liability for Artificial Intelligence and other emerging digital technologies* (European Commission, 2019) contains several references to the application of the technologically neutral principle. When using an AI system during the execution of a contract, there is a high opacity and complexity when gauging whether the activity of the programmer has been successfully considered or not in the result executed by the algorithm. In other words, because of the relatively autonomous nature of artificial intelligence systems, it is not always easy to determine whether the person who "built" the algorithm has properly fulfilled or not their contractual obligations.

As the overall functioning of the system is not always subject to human supervision, it is difficult to determine the contractual liability of the parties. It is not even clear whether any party could have influenced the calculation of the algorithm. On 16 February 2017 the European Parliament adopted a resolution promoting common European definitions for "cyber physical systems, autonomous systems, smart autonomous robots and their subcategories", as well as the possible establishment of a legal personality specific to artificial intelligence systems in order to ensure the existence of liability and the corresponding compensation, in case of damage caused by an artificial intelligence system and not attributable to a human factor.

However, most legal orders have not yet adopted specific rules for regulating the possibility of damage caused by an artificial intelligence system. Thus, when regulating legal responsibility of AI systems, a choice between neutral and AI-specific approaches must be made. It would clearly make sense to try to produce regulations on the legal liability of AI systems in accordance with the principle of technological neutrality; if it is possible to draft the regulation in such terms. However, some authors consider that with a complete technological neutrality, “it is impossible to draft legislation with sufficient precision and clarity that addresses every possible future technical variation” (Bennett Moses, 2005, p. 578).

Most legal orders have not yet adopted specific rules for regulating the possibility of damage caused by an artificial intelligence system. Thus, when regulating legal responsibility of AI systems, a choice between neutral and AI-specific approaches must be made.

An important factor that helps determine whether technologically neutral drafting may be possible or not could be the regulator’s understanding of the technology in question. Otherwise complex schemes of “AI facts” will arise. If member states decide therefore to maintain a technologically neutral approach based on existing liability regimes, they will be quickly overtaken by technological developments in AI. In France, for example, liability regimes for defective products have been conceived from a purely corporal perspective (Articles 1384 and 1386-5 of the French Civil Code).

Therefore, even if they can be applied to AI, they will probably lead to inappropriate and unbalanced solutions, either because they will provide too automatic a responsibility or because they will organise an unfair distribution of responsibilities. Not to mention the problems of legal uncertainty and of harmonised application of technologically neutral regulation by national judges. As for the latter, the national courts may resist technologically neutral laws, because as time passes and technology progresses, it is more difficult to see whether the law should apply or not, leading to judicial uncertainty on law application (Greenberg, 2016). Judges will not be able or willing to apply technologically neutral laws in an equivalent and consistent manner.

The establishment of legal responsibility at EU level set against national regimes illustrates the complexity and risks of a technologically neutral approach. This is why it is important to rethink the concept of technological neutrality, so that future AI regulation is more specific to this emerging technology.

The “Technology-Specific Neutrality” Approach – an Alternative Choice

“Technological neutrality is the dual concern of avoiding conferring a monopoly on a particular commercial technology or product, and also of avoiding freezing the law in relation to a transitional state of technology” (Sorieul, 2003, p. 409). Announcing an ambitious digital strategy for AI and data on 19 February 2020, the European Commission (2020a) has fixed the objectives for the digital future of the Union. It aims indeed to establish a strict legal framework regulating new technologies and effectively controlling innovations. However, the diversity in the application of emerging digital technologies entails a wide range of risks, making it difficult to find simple solutions.

Perceived by many as an advantageous method of legislative drafting, it was instead denounced by Professor Gautrais (2012) for the vague and inclusive character that limits its usefulness. As he points out, it can result in legislative vagueness or hasty reaction by a legislator who has not adequately weighed the interests of all. Technologically neutral regulations give regulators flexibility, but this could encourage them to prematurely extend their authority to new markets and technologies before it is proven that a lasting market failure must be corrected (REFIT Platform Opinion, 2017). In this sense, technological neutrality could encourage over-regulation of new emerging markets.

However, the choice of a legislative technique does not necessarily have to be binary.

Technologically neutral regulations give regulators flexibility, but this could encourage them to prematurely extend their authority to new markets and technologies before it is proven that a lasting market failure must be corrected.

Laws in fact generally result from the combination of neutrality and specificity, at least conceptually. Several considerations are relevant for evaluating legislation and choosing either a technology-neutral or technology-specific approach. If the considerations apply, they become justifications for the chosen technique. In other words, the upcoming European regulation on AI systems will have to observe, in my opinion, three main considerations: flexibility, innovation, harmonisation.

A technologically neutral legislative technique is flexible, as it covers a wide range of technologies, while a law that mentions a specific technology is set to become obsolete sooner or later (Craig, 2013). However, technology-specific legislation is almost by definition more precise and specifically adapted to the problem that the legislation aims to solve. It can identify the problems involved in a particular context (Ohm, 2010, pp. 1695–1696).

Future EU regulation policy on AI systems will have to transcend all economic sectors of the European single market: energy, aviation, health, finance, etc. Each sector's specificities will have to be then taken into consideration when drafting upcoming EU policies on AI systems. For example, in their study on cost-effectiveness support schemes for electricity generation from renewable energy (RES-E), Paul Lehmann and Patrik Söderholm (2015, p. 14) show that "Overall, technology-specific support schemes may thus produce economic benefits, particularly if technology markets work imperfectly and in second-best settings with additional uncorrected market failures." Thus, even if a technology-specific legislation is less flexible, we might prefer sometimes specific and narrowly adapted regulations.

Over the past three years, EU funding for AI research and innovation has increased by 70% compared to the previous period (Gagliani, 2020). A technologically neutral legislation helps foster innovation and avoids problems of law circumvention. On the other hand, the open nature of technologically neutral legislation could act as a deterrent to technology developers. Not knowing in advance how the law could deal the new technology developed, developers might refrain from pursuing it, perhaps to the detriment of all, while specific legislation can provide certainty.

A third consideration that supports technologically neutral legislation is related to chances of achieving harmonisation between different jurisdictions. This consideration applies when the technology, or its use, is not confined to a defined locality and when different jurisdictions are dealing with the same problem more or less simultaneously. Thus, this consideration is obviously relevant for the digital online environment. The case of the eIDAS regulation (electronic IDentification, Authentication, and trust Services), which guarantees that electronic interactions between companies are safer, faster, and more efficient, regardless of the European country in which they take place, illustrates this point. Online transactions often take place across borders. If each state had chosen a particular technology and incorporated it into their legislation through a technology-specific approach, interstate and international barriers to electronic commerce would likely have been raised. Technologically neutral legislation would validate many technologies, not on the basis of their specificities but of their function, and thus allow harmonisation. The downside of going for a broader statutory language is that using many different technologies could lead to a lack of interoperability.

Making EU regulations on AI systems more technology-specific would significantly reduce legal uncertainty. Following Greenberg's research (2016, p. 1547) on the US Copyright Act, which openly abandons the approach of technological neutrality to rather embrace the concept of technological discrimination. The laws should prescribe a neutral treatment of all technologies that fall within an area defined by certain characteristics and specifications (domain-specific neutrality approach). This technological discrimination should thus help to circumvent the unintended consequences of technological neutrality. Such an approach would thus allow a non-legislative adaptation of law to future technologies and would recognise when new technologies should be treated differently from older technologies.

Making EU regulations on AI systems more technology-specific would significantly reduce legal uncertainty.

Conclusion

The future for sustainable development and security in the EU will largely depend on how digital technologies such as 5G, cloud, or AI are implemented. Artificial intelligence, whose evolution is at its peak today, outlines a different future for the socio-economic reality of the modern world. A new area of relatively unregulated social relations has been created, for which specific rules of conduct should be established.

The employment of AI systems is characterised by relative opacity, in the sense that it is not easy to understand how they operate and make decisions. From autonomous weapons systems (AWS) to facial recognition technology to decision-making algorithms, the dual nature of artificial intelligence technology brings enormous security risks to both individuals and entities across nations.

While the debate on the structure, role, and dual use of AI will continue in the coming years, any attempt to redefine AI security needs to begin with identifying, understanding, incorporating, and broadening the definition and nature of AI security threats. So, we must ensure that the digital transformation of our society and our industries is successfully carried out to build a secure digital European market. Therefore, the digital transformation of the Union will not only have to be accompanied by legislative drafting with specific neutrality in the field of Artificial Intelligence, but also by a fully operational strategic autonomy. ■

About the author:

Theodoros Karathanasis

Hellenic national, I acquired a strong experience on management after having been working for seven years in private sector (2005-2012). Since 2017, I hold a double degree in law (Paris II Panthéon Assas, 2016) and economics & management (Univ. of Strasbourg, 2017). While graduating from Sciences Po Grenoble in 2018 on European Governance studies (Master), I completed the same year an internship to the Permanent Representation of Greece to the EU, in the department for relations with the European Parliament, the CoR and the EESC. Today Phd candidate in EU law at my 3rd year at the Grenoble laboratory CESICE (Centre Européen de Sécurité Internationale et des Coopérations Européennes). I obtained a scholarship by the Grenoble Alpes Institute of Cybersecurity to write a thesis on "The effectiveness of the NIS Directive: national actors in the face of European rules on cybersecurity". Therein, a communication was made in 2019 at the AFEE Study Day in Grenoble on "The acceptance of EU law limited by state adaptation: the case of the NIS Directive". With a vivid interest in the concept of "hybrid threats", I try to combine it with my field of study themes related to Artificial Intelligence, the European mechanisms of crisis management (e.g. IPCR, CRS, Argus), the external action of the Union (e.g. CFSP-CSDP) or the instrumentalization of international law (e.g. Tallin Manual 2.0). At the moment, a study on "Assessing Union's Cybersecurity policy 'softness' beyond EU law compliance" is being drafted.

References

Ali, R. (2009). Technological Neutrality. *Lex Electronica*, *14*(2), 1–15. www.lex-electronica.org/en/s/443

Bennett Moses, L. (2005). Understanding Legal Responses to Technological Change: The Example of In Vitro Fertilization. *Minnesota Journal of Law, Science and Technology*, *6*(2), 505–618.

Caresche, C. (2012). *Rapport fait au nom de la Commission des finances, de l'économie générale et du contrôle budgétaire, sur le projet de loi portant diverses dispositions d'adaptation de la législation au droit de l'Union européenne en matière économique et financière (n° 232)*. https://www.assemblee-nationale.fr/14/pdf/rapports/r0469.pdf

Copyright Law of the United States and Related Laws Contained in Title 17 of the United States. (June 2020). https://www.copyright.gov/title17/title17.pdf

Craig, C. J. (2013). Chapter 9 - Technological Neutrality: (Pre)Serving the Purposes of Copyright Law. In: Geist, M. (Eds.). In: *The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law*. Les Presses de l'Université d'Ottawa | University of Ottawa Press.

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L1024

Directive 2002/21/EC of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive). https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32002L0021

Directive 2009/140/EC of the European Parliament and of the Council of 25 November 2009 amending Directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and associated facilities, and 2002/20/EC on the authorisation of electronic communications networks and services. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32009L0140

European Commission. (2019). *Liability for Artificial Intelligence and Other Emerging Digital Technologies*. https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608

European Commission. (2020a). Shaping Europe's Digital Future. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_273

European Commission. (2020b). White Paper: On Artificial Intelligence - A European approach to excellence and trust. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52017IP0051

European Parliament resolution of 12 February 2020 on automated decision-making processes: ensuring consumer protection and free movement of goods and services. Retrieved from: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.html

European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies. https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.html

Gagliani, G. (2020). Cybersecurity, Technological Neutrality, and International Trade Law, *Journal of International Economic Law*, *23*(3), 723–745.

Gautrais, V. (2012). *Neutrality Technological: drafting and interpreting laws in the face of technology*, Themis, Montreal.

Greenberg, B. A. (2016). Rethinking Technology Neutrality. *Minnesota Law Review*, 207, 1495–1562.

High-Level Expert Group on Artificial Intelligence. (2019). A definition of AI: Main capabilities and scientific disciplines. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

Karanasiou, A., & Pinotsis, D. (2017). *Towards a Legal Definition of Machine Intelligence: The Argument for Artificial Personhood in the Age of Deep Learning*. International Conference on Artificial Intelligence - Law (ICAIL).

Koops, B.-J. (2006). Should ICT Regulation be Technology Neutral? In: B.-J. Koops, M. Lips, C. Prins Mr. Schellekens (Eds.). *Starting Points for ICT Regulation. Deconstructing Prevalent Policy One-Liners*. The Hague: T.M.C. Asser Press, pp. 77–108.

Lehmann, P., & Söderholm, P. (2015). Technology-neutral or Technology-specific? Designing Support Schemes for Renewable Energies Cost-effectively. International Association for Energy Economics, Antalya Special Issue, 13–15.

Lozano, J., & Murillo, J. (2020). What should be taken into account if Artificial Intelligence is to be regulated? BBVA. https://www.bbva.com/en/opinion/what-should-be-taken-into-account-if-artificial-intelligence-is-to-be-regulated/

Maxwell, W. J., & Bourreau, M. (2014). Technology Neutrality in Internet, Telecoms and Data Protection Regulation. *Computer and Telecommunications Law Review*, *21*(1), 1–4.

Ohm, P. (2010). The Argument against Technology-Neutral Surveillance Laws. *Texas Law Review*, *88*(7), 1685–1714.

REFIT Platform Opinion on the submission by the Royal Norwegian Ministry of Trade, Industry and Fisheries on Intention, Digitalisation and Technology Neutrality. (23 November 2017). https://ec.europa.eu/info/sites/info/files/xxii-3b-technology-neutrality_en.pdf

Regulation (EU) 2015/2120 of the European Parliament and of the Council of 25 November 2015 laying down measures concerning open internet access and amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services and Regulation (EU) No 531/2012 on roaming on public mobile communications networks within the Union (Text with EEA relevance). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32015R2120

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018R1807

Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R0881

Sorieul, R. (2003). The CNUDCI Electronic Signatures Model Act. In: Chatillon, G. (Dir.), International Internet Law: Acts of the symposium organized in Paris on 19 and 20 November 2001 by the Ministry of Justice, the University of Paris I Pantheon Sorbonne and the Association Arpeje, Brussels, Bruylant.

The Hague District Court. (2020). Case number / cause list number: C/09/550982 / HA ZA 18-388. Judgment of 5 February 2020. https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878

The International Bank for Reconstruction and Development, The World Bank, InfoDev & The International Telecommunication Union. (2011). *Telecommunications Regulation Handbook*. https://www.itu.int/dms_pub/itu-d/opb/pref/D-PREF-TRH.1-2011-PDF-E.pdf

ANALYSIS

Integrated IT-OT Assessment and Governance Model for Improved Holistic Cybersecurity

Dr. ENRICO FRUMENTO

SENIOR DOMAIN EXPERT, CEFRIEL - POLITECNICO DI MILANO

ABSTRACT:

The Operational Technology (OT) is a novel and rapidly expanding area for both cybercrime and industry. The number of attacks against OT infrastructures is increasing; the pandemic played a considerable role because of the digital transformation's acceleration. For example, one of the impacts of COVID-19 is the reduction of on-site staff, which puts a strain on OT systems, on the already limited resources and required an increase in external connectivity. However, from a cybersecurity point of view, IT and OT are still missing a holistic approach that includes cybersecurity, physical, and cyber-physical security, an integrated cyber-risk estimation, and governance models able to span across IT OT domains.

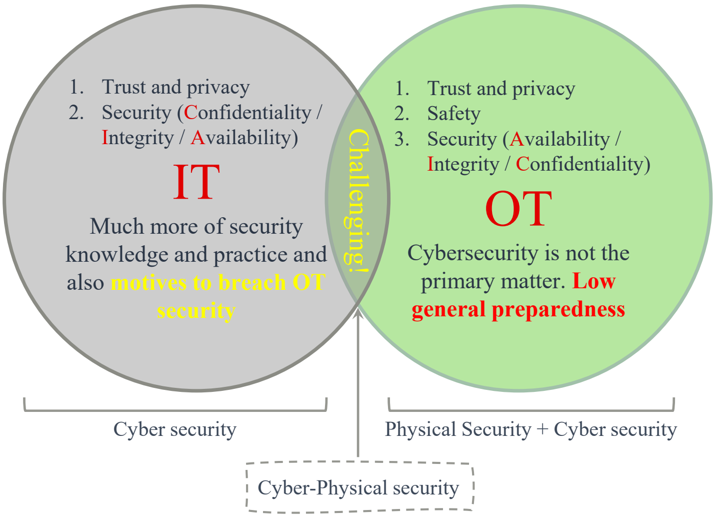
Keywords: Operational Technology, cybersecurity, IT-OT convergence

Cybersecurity: Incorporate Data Integrity Strategies

Gartner defines Operational Technology (OT) as “hardware and software that detects or causes a change, through the direct monitoring and/or control of industrial equipment, assets, processes and events”. OT differs from IT in terms of functionalities, the culture of operators, and threats.

IBM’s 2020 X-Force Threat Intelligence Index summarises that attacks targeting operational technology (OT) infrastructure increased by over 2,000 per cent in 2019 compared to the previous year (Kovacs, Echobot Malware Drives Significant Increase in OT Attacks, 2020). Security requirements of CPSs (Cyber-Physical Systems) are growing in parallel to the evolution of their threat landscape. CPS security goes under the umbrella of the Operational Security (OT), a branch of computer security that differs from IT security in several respects. IT Security typically builds up from the Confidentiality-Integrity-Availability paradigm (CIA), while OT Cybersecurity starts from the Safety-Reliability-Productivity (SRP) properties. Hence, the safety and security aspects in the CPSs are tightly connected to each other. Ten years ago, OT systems were physically separated from IT systems and the threat environment was limited. Today, we witness a convergence of IT and OT systems instead: protecting modern CPS installations requires both information technology (IT) and operational technology (OT) expertise (Gary & Prananto, 2017; D., 2018). Gartner, in its hype cycle for the Internet of Things 2019, reports the IT/OT alignment at the beginning of the plateau of productivity (Gartner, 2019); however, as shown in Figure 1, there is a challenging area that Gartner chart does not consider, being the hype cycle about IoT and not about IIoT and CPSs. Recent literature reports that the cybersecurity approach must be holistic, including cybersecurity, physical security, and cyber-physical security (see below). Its governance model must be the same – spanning across IT and OT domains (Benias & Markopoulos, 2017). This is a challenging area still in the focus of research. Moreover, the COVID-19 pandemic hastened this trend: it is the digital accelerant of the decade, speeding up companies’ digital transformations worldwide by approximately an average of six years. This category of problems first gained momentum with the case of Norsk Hydro (Fouche & Solsvik, 2019), where an IT attack provoked OT consequences that rolled into the company up to the governance level, which took the decision to stop the production line (Kovacs, Industry Reactions to Norsk Hydro Breach: Feedback Friday, 2019).

Figure 1. Areas of IT-OT security. Source: Ghaznavi, 2017.


1. Trust and privacy
2. Security (Confidentiality / Integrity / Availability)
IT
Much more of security knowledge and practice and also motives to breach OT security
Challenging!
1. Trust and privacy
2. Safety
3. Security (Availability / Integrity / Confidentiality)
OT
Cybersecurity is not the primary matter. Low general preparedness
Cyber security
Physical Security + Cyber security
Cyber-Physical security

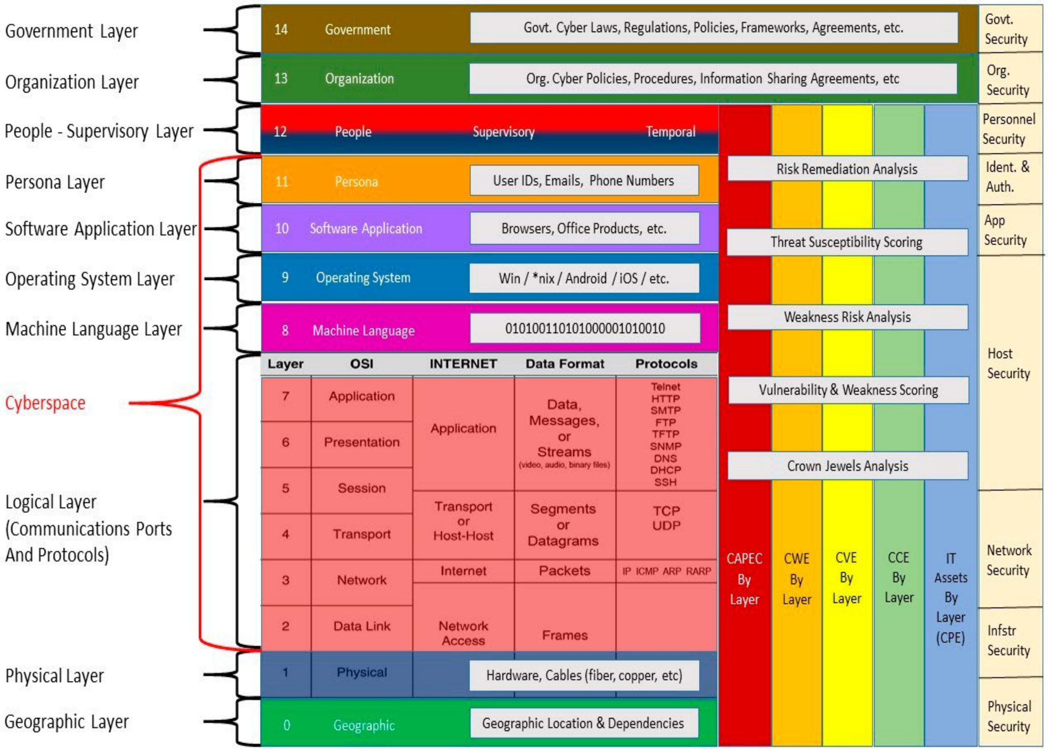
Cybersecurity approach must be holistic, including cybersecurity, physical security, and cyber-physical security, and its governance model must be the same – spanning across IT and OT domains.

The most critical context where these problems are present today is data-intensive Industry 4.0, where data collected, usually at the edge, are passed along a relatively complex chain of technological handlers. In this context, cybersecurity is necessary. Unfortunately, complex data-intensive scenarios, like those requested by modern Industry 4.0, are difficult to govern, also because of the CIA and SRP criteria mentioned above. However, literature does not agree on what is the correct model for a correct holistic (i.e. omni-comprehensive) approach to cybersecurity (an example is the VMWare's Cognitive attack Loop (VMware Carbon Black, 2019)). Putting together all the layers that are affected by a cyber threat leads to a relatively complex model, made of fifteen layers, which starts from the root (the physical layers) and ends with integrated governance. The so-called Cyber-Terrain Model or CTM (Riley, 2014) originated in the military area (David, Conti, Cross, & Nowatkowski, 2014) and is gaining momentum as one of the most complete models for cybersecurity in mixed IT-OT data-intensive contexts.

Regarding Figure 2, the CTM:

- can enable a shared understanding by engineers, operators, analysts, executives, and board members, which led to the adaptation of the 15-layer model shown below to the OT context;
- has clear links with attacker TTPs (Tactics, Threats, and Procedures), which are fundamental information to guide the vulnerability assessment;
- supports the extension of Vulnerability Assessments over its 15 layers.

Figure 2. Cyber-Terrain Model layers. Source: Riley, 2014.



Another crucial distinction is the difference between the IT and OT threat landscapes (Fink & McKenzie, 2019): the risks and threats of these two worlds are very different. For example, one relevant ICS threat is the miss or delay of required traffic: while in an IT system this would not be in most cases a related issue, in a CPS it becomes a threat because of the anomalies injected into the system (or the malicious absence of anomalies, e.g. an infected and malfunctioning system that reports normal logs).[1] By this point of view, literature reports that anomaly detection solutions for a CPS should differentiate, and handle differently, real anomalies (i.e. coming from defective processes or actual exceptional events) and injected anomalies (i.e. resulting from a cyber-attack). Since Stuxnet (Denning, 2012), OT security has been demonstrating that malware can inject anomalies or drift in the processes. Unfortunately, cybersecurity in the CPSs and Industry 4.0, in general, still have low maturity. Regulations such as ISA99/IEC 62443 (Network and system security for industrial-process measurement and control) are still not complete – for example, Part 3-2 (Security risk assessment and system design and technical requirements) is in draft.

With the emergence of the Internet of Things paradigm (and its contextualisation to the Industrial IoT setting), the ability to connect automated devices that communicate via the Internet is becoming pervasive from the factory floor (Lezzi, Lazoi, & Corallo, 2018). The transition from closed networks to enterprise IT networks and then to the Internet is increasing issues and alarms about security. As we increasingly rely on intelligent and interconnected devices, a new question related to security comes up: how can we protect all the appliances to avoid the intrusions and interferences that could compromise personal security and privacy? The confidence in these devices has become essential, and it is a crucial factor to guarantee cybersecurity. Vulnerabilities are usually based on software failures used to force the device to change its normal behaviour or operation (Radanliev, et al., 2018). These vulnerabilities are intrinsic to the software, but it is possible to reduce them with a good design and software implementation. The main reason for this is that the software is usually handmade, and errors remain in the final code.

As we increasingly rely on intelligent and interconnected devices, a new question related to security comes up: how can we protect all the appliances to avoid the intrusions and interferences that could compromise personal security and privacy?

Another research question is related to IoT devices' connectivity **(or rather Industrial IoT)** or new software and older applications. This type of applications is usual in the industry, like embedded programmable controllers and *automata* operating systems that sometimes are integrated into enterprise IT infrastructure. In this sense, it is vital to protect them from human interference while preserving the investment in the IT infrastructure and the leverage on the security functions. Besides, the way these systems receive software updates and patches without risk in terms of safety is remarkable. Moreover, the channels must be secure to protect the information from unauthorised disclosure and usage. On the other hand, the use of insecure peripherals can, therefore, infect the network with malware, leading to potential dangers when reaching the production control system.

Existing techniques in the state of the art for the domain of high-reliability systems must be used efficiently in the designing, developing, and maintaining phases of these new connected applications. Security must play a vital role in the whole

[1] Two significant categories make up the threat landscape of CPS: (i) malevolent agents that are injecting anomalies and (ii) malevolent agents that are injecting normalities. The type (i) refers to agents that, more or less rapidly, shift the operational parameters of the CPS or the entire production floor into the direction planned by the attacking entity (e.g. malware which deliberately and slowly alters a turbine speed or an oven's temperature). The type (ii) refers to agents that, controlled by the attacker, are reporting normal behaviour to the monitoring sensors in the face of malfunction, to mask their activities.

data life cycle and throughout the operational environment (Ustundag & Cevikcan, 2017). Some of these security techniques are related to:

- Cryptographically generated signatures. They are usually used to create whitelists of users that minimise the risks of contaminating the system through other applications and patches.
- Device authentication: when a new device is plugged into the network, it should authenticate itself before receiving or transmitting data. With machine authentication, the devices can access a system based on credentials stored in a secure area.
- Updates and patches: when a device is working, it will receive software updates similar to a computer with Linus or Windows operating system, with the security problems and faults corrected as soon as they are known. Of course, it is a problem to patch thousands of devices which are executing critical functions with high availability constraints. These patches must also preserve limited bandwidth and connectivity as well as use secure transmission links.
- Hypervisors: to protect the applications with several levels of security, a hypervisor allows running different operating systems side by side within compartments whose strict separation is ensured by hardware support. Of course, it is another problem to implement virtualisation on low-end computing systems with limited computing capabilities.
- Firewalls and IPS: devices need a firewall or some other network data inspection capability to control incoming and outgoing traffic with other devices. Some of the industrial embedded devices do not use IP based protocols, but instead they use specific protocols such as MODBUS (Grid Connect, 2019). Thus, industry-specific protocols also need filtering and malware inspection capabilities.

Security must play a vital role in the whole data life cycle and throughout the operational environment.

Data security must be guaranteed along the entire data life cycle: at ingestion or acquisition time (where the data is generated), at transportation or motion stage (while moving from sensors to the destination), and at rest phase (once stored, for example in the database)[2].

Impact of 5G and the Evolution to Industry 5.0

According to McCann et al., “The Fourth Industrial Revolution or simply ‘Industry 4.0’ is how the manufacturing industry expects to maximise the innovations of 5G wireless communications by automating industrial technologies and utilising other enabling technologies such as artificial intelligence (AI) and machine learning. The industry expects this to lead to more accurate decision making such as automation of physical tasks based on historical information and knowledge, or improved outcomes for a wide range of vertical marketplaces not just in manufacturing but verticals such as agriculture, supply chain logistics, healthcare, energy management and an ever-increasing number of industries becoming more aware of the potentials of 5G” (McCann, Quinn, McGrath, & O’Connell, 2018).

While 4G networks can provide peak data rates of one gigabit per second (Gbps) and actually sustained bandwidth of around 10 Mbps, new industry services will significantly exceed what current networks can reliably provide. Within 5G, the experienced bandwidth is expected to increase 100-fold and reach up to one Gbps. Frequently, applications also have intense demands on data delay. For instance, in autonomous driving or critical control systems, when data must be shared across devices, latencies of one millisecond are a target, which is a tenth of what is provided by current mobile networks.

[2] The three phases of the data life cycle are (1) ingestion (acquisition from local sensors), (2) data in transit and (3) data at rest (Lord, 2019).

Adopting 5G in Industry 4.0 has already started with installations of private 5G networks;[3] however, todays incoming large-scale adoption of 5G started to be referenced as the transition from Industry 4.0 to Industry 5.0. Industry 5.0 will be characterised by the cooperation between machines and human beings, with the aim to give an added value to production, by creating personalised products able to meet customers' requirements (Madia, 2020). Industry 5.0 is of paramount importance for Europe. 5G can become the reference communication platform for almost all industrial sectors, driving the transition of Industry 4.0 and intelligent manufacturing towards Industry 5.0. In this sense, 5G becomes an element of the supply chain and the industrial threat landscape (O'Connell, Moore, & Newe, 2020). Unfortunately, it happens in a complex worldwide context, shaken by pandemics, cybercrime raise, economic tensions and transformations.

- Being the 5G part of the OT threat landscape feeds cyber threats and endangers the safety and opens potential cyberwarfare scenarios.
- Private networks are not subject to EU Toolbox and can make their installations without legal limitations, nor are they subject to Golden Power[4] requirements. From an industry point of view, a private 5G network opens the possibility of adding layered security at every level of the manufacturing process. This affords the network owner a level of control that would not be possible on public infrastructure.
- Safety and cyber warfare scenarios are not theoretical. However, even in less drastic scenarios, following a 5G slowdown, the control chains might suffer hiccups with consequences for an industry's control layer (e.g. protection against industrial espionage).
- 5G must guarantee ultra-reliable, high-speed, low-latency, power-efficient, high-density wireless connectivity and safety and reliability in the OT sense, 99.9 periodic per cent of the time.
- 5G enables the exploitation of processing power and storage capabilities that cloud servers boast, thus representing a shift towards the Integrity-Availability-Confidentiality (IAC) paradigm. Therefore, it is crucial to keep in mind the network segmentation and guarantee the integrity and availability of data. Moreover, there is a need to handle the concurrent access to data properly while maintaining the established service level agreement.

> **5G can become the reference communication platform for almost all industrial sectors, driving the transition of Industry 4.0 and intelligent manufacturing towards Industry 5.0. In this sense, 5G becomes an element of the supply chain and of the industrial threat landscape.**

Suppose 5G-controlled factories allow an attacker to compromise the integrity of production. In that case, such a tool will destroy or hurt a competitor much more elegantly than a brute-force disruption of the production. Alternatively, from the defender's viewpoint: it is good to know that the machines are running, but it is even more important to know that they are producing what they are supposed to produce.

[3] A mobile private 5G network, as the name suggests, is a local area network that utilizes 5G technology as its communication medium to build and create a "private" network. Private 5G networks are referred to as local 5G networks or mobile private networks (MPN). This involves two options: (i) the deployment of a physically isolated private 5G network (i.e. a 5G island) that is independent of the mobile operator's public 5G network; (ii) the deployment of a private mixed mode 5G/4G network which connects to a mobile operator through a 4G or internet connection.

[4] Italy, like France with the "Bothorel" law, has a special legislation for critical infrastructures of national interest, called "Golden Power". It is a national regulation that defines rules for acquirers of technologies from non-European vendors. The legislation was extended to include the 5G networks and operators (Article 1-bis of L 21/2012, special powers relating to broadband electronic telecommunications networks with 5G technology). This is an Italian set of obligations, with specific extensions for 5G, that applies to all the relevant national security actors (e.g. MNOs, public bodies, critical infrastructures).

Recent SwissRe SONAR Insurance report of 2019 (SwissRE, 2019) reinforce the discussion: *"hackers can also exploit 5G speed and volume, meaning that more data can be stolen much quicker [...] interruption and subversion of the 5G platform could trigger catastrophic cumulative damage. Cyber exposures are significantly increased with 5G, as attacks become faster and higher in volume"*.

Integrated IT-OT Protection Solution

Summing OT security, from a bird's-eye view, has some main problematic areas:

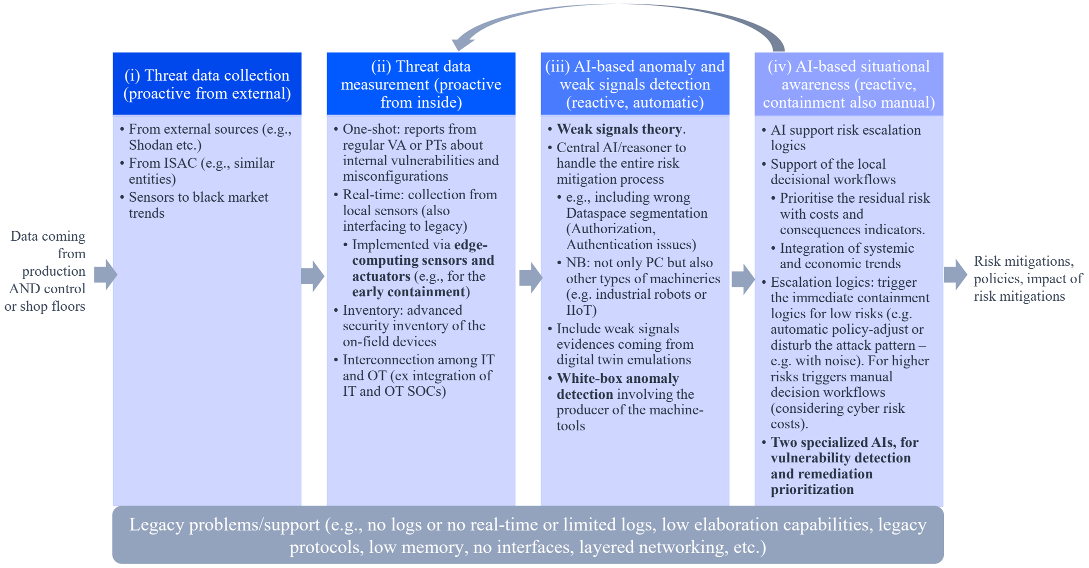
- **Protect boundaries.** OT machines have a long operation life (usually up to 30 years) and often cannot be live-patched. It is therefore often impossible to protect individual devices against cyber-attacks. Consequently, it is necessary to protect the boundaries.
- **OT is in the middle of the supply-chain (island hopping).** The most common OT risk is being someone else's supply-chain "Trojan" (one of the islands in an "island hopping" attack, e.g. Siemens devices with StuxNet).
- **Very long and usually low controllable supply-chains, also for relatively small industries and impact of intangible assets**. The loss of reputation (i.e. an intangible asset) in front of the final client after a critical data breach is one of the most neglected risks. This risk requires careful selection of the supply chain members, those involved in the maintenance process, high-security standards and third-party liability clauses.
- **Low maturity of cybersecurity in Industry 4.0/5.0.** Cyber-security maturity of I40 is still low compared to IT. Fundamental regulations such as IEC 62443 (Network and system security for industrial-process measurement and control) are always evolving in cybersecurity chapters. The market is rapidly growing and researches too. This happens despite Gartner's hype cycle on Internet of Things for 2019 (Memon, Memon, Ahmed, Ahmed, & Sattar, 2020) reporting IT/OT convergence in the "Slope of Enlightenment" phase while the 2020 edition phase it out of the hype cycle.

As reported in Figure 3, a schematic approach that encompasses all the considered remarks is made of four phases: (i) threat data collection, (ii) threat data measurement, (iii) AI-based anomaly and weak signals detection, and (iv) AI-based situational awareness. The first two are related to the proactive collection of attack evidence from outside (e.g. proactive threat intelligence scanning) and from inside (e.g. proactive vulnerability assessments) an OT implant. The phases (iii) and (iv), in turn, are two AI-enabled phases where the information is collected and analysed to identify inconsistencies and anomalies.

Figure 3 represents a dynamic threat management solution made of the following elements:

- (i) Data of externally known, incoming threats fed into the system (e.g. Shodan, direct scanning and dark web, honeypots, malware, SSL to track actors across the Internet, big data sharing projects, black market, etc.).
    - Comparison with similar entities (e.g. other industries) to identify the incoming threats and do data analysis on the collected information.
    - Threat intelligence system solutions and input from external entities such as ISAC or CSIRT, or known vulnerability databases such as CVE, CWE, and NVD. In addition to them, MITRE's ATT&CK and CAPEC are two frameworks that collect attack methodologies used to exploit component vulnerabilities (Mavroeidis & Bromander, 2017). The information obtained can be represented in machine-readable formats like JSON (on which the STIX standard language for cyberthreat intelligence information exchange is based) or CVRF (with which the vulnerabilities from CVE database are represented) (Ramsdale, Shiaeles, & Kolokotronis, 2020).

Figure 3. General schema of the proposed dynamic threat management solution for OT systems.



Data coming from production AND control or shop floors

(i) Threat data collection (proactive from external)

- From external sources (e.g., Shodan etc.)
- From ISAC (e.g., similar entities)
- Sensors to black market trends

(ii) Threat data measurement (proactive from inside)

- One-shot: reports from regular VA or PTs about internal vulnerabilities and misconfigurations
- Real-time: collection from local sensors (also interfacing to legacy)
  - Implemented via **edge-computing sensors and actuators** (e.g., for the **early containment**)
- Inventory: advanced security inventory of the on-field devices
- Interconnection among IT and OT (ex integration of IT and OT SOCs)

(iii) AI-based anomaly and weak signals detection (reactive, automatic)

- **Weak signals theory**.
- Central AI/reasoner to handle the entire risk mitigation process
  - e.g., including wrong Dataspace segmentation (Authorization, Authentication issues)
  - NB: not only PC but also other types of machineries (e.g. industrial robots or IIoT)
- Include weak signals evidences coming from digital twin emulations
- **White-box anomaly detection** involving the producer of the machine-tools

(iv) AI-based situational awareness (reactive, containment also manual)

- AI support risk escalation logics
- Support of the local decisional workflows
  - Prioritise the residual risk with costs and consequences indicators.
  - Integration of systemic and economic trends
- Escalation logics: trigger the immediate containment logics for low risks (e.g. automatic policy-adjust or disturb the attack pattern – e.g. with noise). For higher risks triggers manual decision workflows (considering cyber risk costs).
- **Two specialized AIs, for vulnerability detection and remediation prioritization**

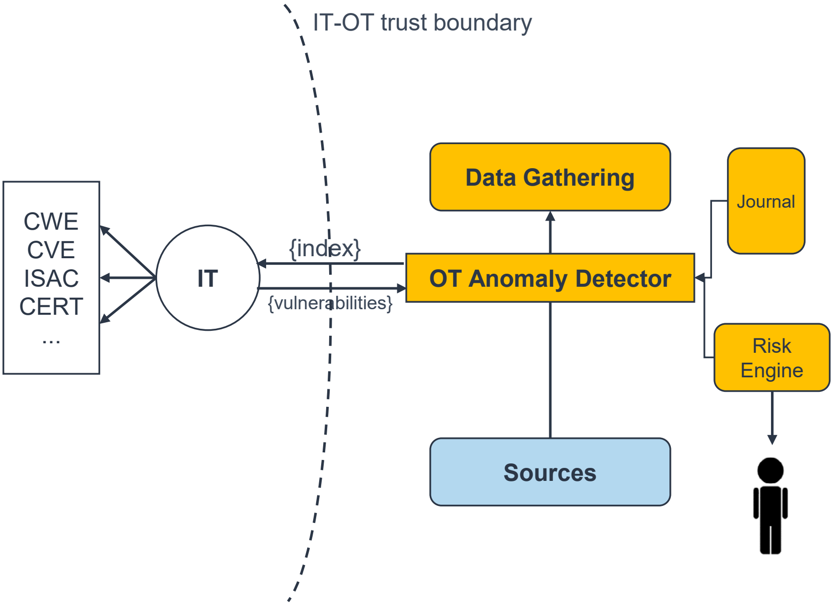
Risk mitigations, policies, impact of risk mitigations

Legacy problems/support (e.g., no logs or no real-time or limited logs, low elaboration capabilities, legacy protocols, low memory, no interfaces, layered networking, etc.)

- (ii) Local sensors: a distributed sensor network, based on edge computing that collects signals from the production and shop floor and sends them to a central AI that detects the weak signals and the anomaly (the idea comes from anomaly detection theory).
  - Collecting device's firmware-level details, including OS specifics, application version, and serial number, for instance through the Nmap tool that makes available scripts, based on SMTP protocol, with the purpose of extracting information from devices (Karlsson & Mazzurco, n.d.).
- (iii) AI doing vulnerability identification. The AI extracts weak signals and improves the anomaly detection engine.
  - Involvement of the machine tools to do white-box anomaly detection (based on templates) also exploiting the "digital twin" approach.
  - Anomaly detection for the rest of the production line, with templates based on measurements (measured templates).
- (iv) AI doing risk prioritisation. The system feeds another AI that specialises in risk mitigation, matching the risk reduction strategies dynamically.
  - The second AI handles the risk management logics.
  - Low risk is automatically contained (e.g. a new policy is issued).
  - High risk that requires some drastic interventions or cannot be automated (e.g. a temporary shutdown of the production floor) is reported along the decisional workflows with a situational awareness platform to the decision-makers. The decision takers evaluate a manual intervention's pros and cons, also economically (Frumento & Dambra)). The decision is taken to the governance board, to support the key decision-makers (e.g. CISO) (Karlsson & Mazzurco, n.d.).
    - Many reports show that support of the internal decision flows is a business pain.
    - Solve the "fear" of applying a patch to a system to solve a threat (e.g. "it's better to stay vulnerable than go out of business").

Figure 4 reports a simplified logic schema of the workflow of Figure 3. Between the data sources and the data gathering phases of a typical OT control implant, we added an OT Anomaly Detector, which plays a pivotal role between IT and OT. Its function is to collect information from both domains and support the identification of anomalies from either the inside or outside (external) points of view, mentioned in Figure 3 phases (i) and (ii).

Figure 4. Schematic representation of the OT Anomaly detection system.



Besides, IT/OT security must link security and safety (i.e. physical security) workflows. As discussed in earlier sections, in OT, safety and security overlap each other. IT and OT workflows, policies, and operations should overlap too. Unfortunately, most IT/OT solutions approach the problem from a purely cyber or safety perspective. Unique governance spans security and safety, cybersecurity and physical SOCs (e.g. internal cam, anomalies in entrances, physical security alarms, etc.).

The last fundamental element is to support the decision workflows from lower to upper levels. This means to support different types of stakeholders with KPIs specific for each of the 15 abstraction layers of the CTM: this includes stakeholders ranging from the production floor, where the typical KPI is availability, up to the governance, where other variables need to be considered (e.g. the cascading economic impact or risk). An interesting area of exploration is to include research results from different sectors such as the civil protection sector (typically used to deal with safety), economy experts (to evaluate the economic consequences of risk and the sustainability of the countermeasures).

Major Innovations and Conclusions

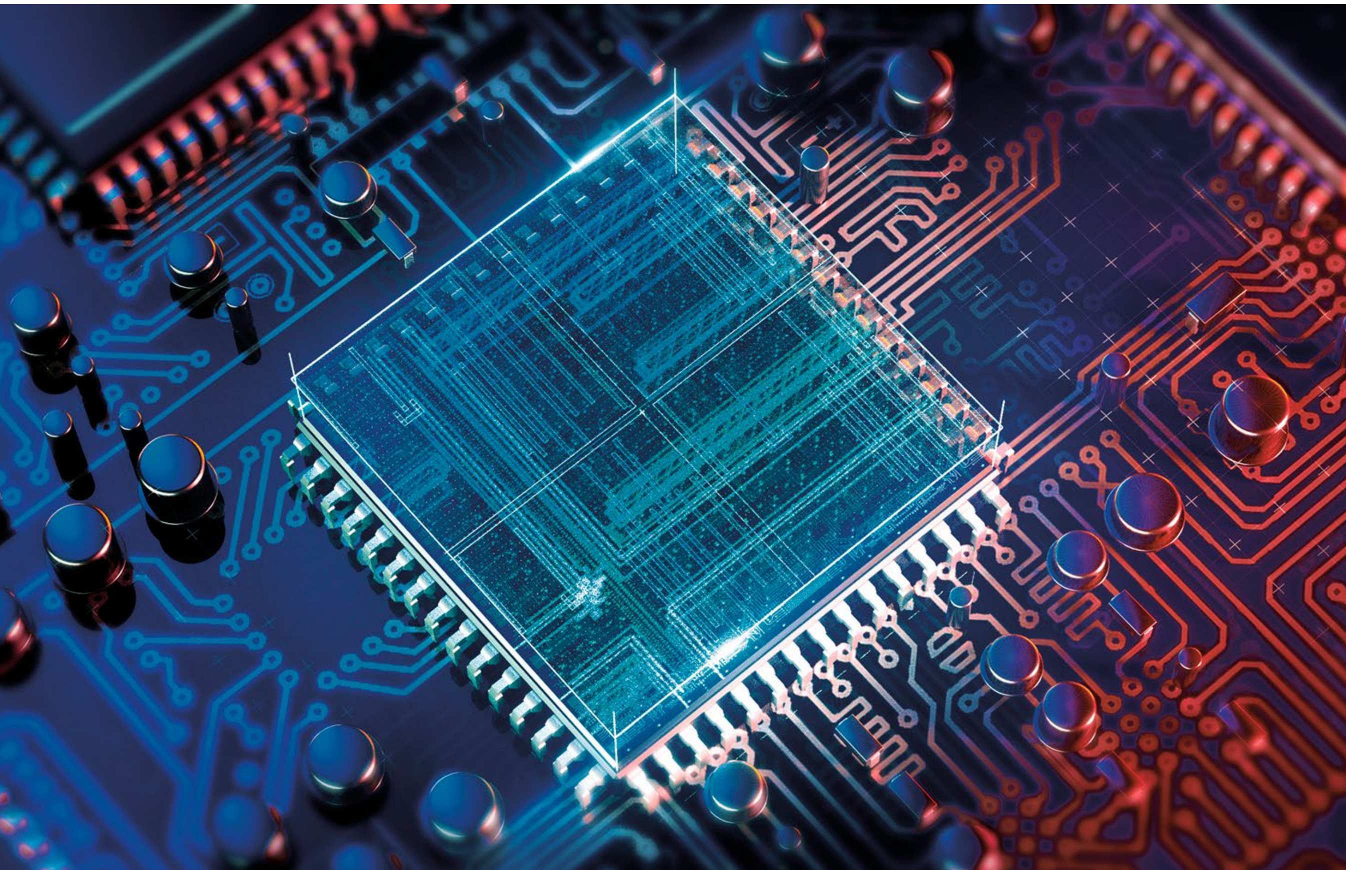
One of the most urgent needs is to **reconcile IT Security** (typically built on Confidentiality-Integrity-Availability paradigm) **with OT Cybersecurity** (built on Safety-Reliability-Productivity properties). This passes through creating a **holistic and sustainable approach** that includes cybersecurity, physical security, and cyber-physical security. The **governance model** must be the same spanning across IT, and OT domains and the **sustainability** must be declined in terms of risks, costs, and processes.

From the cybersecurity point of view, the significant advantage, over state of the art, will be creating an integrated security solution, designed to protect and ensure the integrity of the entire data-processing chain. As it happened already in other sectors, for example, software development, the security countermeasures are considered an add-on, added later in the development stage on top of the data analytics or AI systems. Like in software development earlier, this approach creates problems because the security solutions are disconnected from the real value created by the system they protect.

From the cybersecurity point of view, the significant advantage, over state of the art, will be in creating an integrated security solution, designed to protect and ensure the integrity of the entire data processing chain.

The second element of innovation is in the security solution itself, which is created considering the two most used IT and OT security paradigms: CIA and SRP. In industrial contexts, more than elsewhere, the conjunction of the CIA and SRP paradigms is a fundamental aspect because security threats are a matter of protecting data and safeguarding production processes and potentially lives. Gartner predicts 75% of CEOs will be personally liable for cyber-physical security incidents by 2024 (Gartner, 2020). The security approach plans explicitly to improve and harmonise IT and OT exchange of information to enhance detection of security-injected anomalies. ■

Partially funded by the OK-INSAID (Operational Knowledge from Insights and Analytics on Industrial Data) project, Italian Minister of Economic Development (MiSE), grant no. ARS01_00917.



About the author:

Dr. Enrico Frumento is a Senior Domain Specialist in the cybersecurity team at Cefriel, ICT Center of Excellence for Research, Innovation, Education and industrial Labs partnerships. He is the author of subject-related publications and books and member of the European CyberSecurity Organisation and the European Digital SME Alliance. His 20+ years of research activity focuses on unconventional security, cybercrime intelligence technologies tactics and techniques, the contrast to the modern social engineering and dynamic assessment of organisations' vulnerabilities corresponding to tangible and intangible assets at risk.

References

Benias, N., & Markopoulos, A. (2017). A review on the readiness level and cyber-security challenges in Industry 4.0. *South Eastern European Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM).*

D., E. (2018). *IT+OT Cyber security experts?* Retrieved from https://www.linkedin.com/pulse/itot-cyber-security-experts-daniel-ehrenreich

David, R., Conti, G., Cross, T., & Nowatkowski, M. (2014). Key Terrain in Cyberspace: Seeking the High Ground. *6th International Conference on Cyber Confl ict*. Tallinn.

Denning, D. (2012). Stuxnet: What Has Changed? *Future Internet*, *4*(3), 672-687.

Fink, G., & McKenzie, P. (2019). Helping IT and OT Defenders Collaborate. *ArXiv*. Retrieved from https://arxiv.org/abs/1904.07374v1

Fouche, G., & Solsvik, T. (2019). *Aluminum maker Hydro battles to contain ransomware attack*. (Reuters) Retrieved from https://www.reuters.com/article/us-norsk-hydro-cyber/aluminum-producer-hydro-hit-by-cyber-attack-on-tuesday-idUSKCN1R00NJ

Frumento, E., & Dambra, C. (n.d.). The HERMENEUT Project: Enterprises Intangible Risk Management via Economic Models based on Simulation of Modern Cyber Attacks. *Proceeding of ICISSP 2018*, (pp. pp 495-502). Prague.

Gartner. (2019). *Hype Cycle for the Internet of Things, 2019*. (Gartner) Retrieved from https://www.gartner.com/en/documents/3947474/hype-cycle-for-the-internet-of-things-2019

Gartner. (2020, 09 01). *Gartner Predicts 75% of CEOs Will be Personally Liable for Cyber-Physical Security Incidents by 2024*. Retrieved from https://www.gartner.com/en/newsroom/press-releases/2020-09-01-gartner-predicts-75--of-ceos-will-be-personally-liabl

Gary, A., & Prananto, U. (2017). Cyber Security in the Energy World. *Asian Conference on Energy, Power and Transportation Electrification (ACEPT).*

Ghaznavi, A. (2017). *Cyber-physical System Security in Smart Power Grids*. (Yazd University) Retrieved from https://www.slideshare.net/AhmadrezaGhaznavi/cps-sec-sg-sg2017-confiran-84641279

Grid Connect. (2019). *Industrial Protocols – Grid Connect*. Retrieved from https://www.gridconnect.com/pages/industrial-protocols

Karlsson, P., & Mazzurco, G. (n.d.). *Library SNMP*. (Nmap) Retrieved October 2020, from https://nmap.org/nsedoc/lib/snmp.html#script-args

Kellermann, T. (2019, July 31). *Cognitions of a Cybercriminal, Introducing the Cognitive Attack Loop and the 3 Phases of Cybercriminal Behavior.* (VmWare, Ed.) Retrieved from https://www.carbonblack.com/2019/07/31/introducing-the-cognitive-attack-loop-and-the-3-phases-of-cybercriminal-behavior/

Kovacs, E. (2019). *Industry Reactions to Norsk Hydro Breach: Feedback Friday*. (SecurityWeek) Retrieved from https://www.securityweek.com/industry-reactions-norsk-hydro-breach-feedback-friday

Kovacs, E. (2020, Feb 11). *Echobot Malware Drives Significant Increase in OT Attacks*. (Security Week) Retrieved from https://www.securityweek.com/echobot-malware-drives-significant-increase-ot-attacks

Lezzi, M., Lazoi, M., & Corallo, A. (2018). Cybersecurity for Industry 4.0 in the current literature: A reference framework. *Computers in Industry*, *103*, 97-110.

Lord, N. (2019, 07 15). *Definition of Data In Transit vs. Data At Rest*. Retrieved from https://digitalguardian.com/blog/data-protection-data-in-transit-vs-data-at-rest

Madia, I. (2020). *Industry 5.0: Towards A New Revolution*. (Criticalcase) Retrieved 2020, from https://www.criticalcase.com/blog/industry-5-0-towards-a-new-revolution.html

Mavroeidis, V., & Bromander, S. (2017). Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. *European Intelligence and Security Informatics Conference.* Attica, Greece.

McCann, J., Quinn, L., McGrath, S., & O'Connell, E. (2018). Towards the Distributed Edge–An IoT Review. *12th International Conference on Sensing Technology*, (pp. 263–268). Limerick (IL).

Memon, I., Memon, S., Ahmed, J., Ahmed, R., & Sattar, A. (2020). FLA-IoT: Virtualization Enabled Architecture for Heterogeneous Systems in Internet of Things. *International Journal Of Advanced Computer Science And Applications*, *11*(4). doi:10.14569/ijacsa.2020.0110450

MITRE. (2020). *Introducing the MITRE ATT&CK Framework for Industrial Control Systems.* (Tripwire, Ed.) Retrieved from The State of Security: https://www.tripwire.com/state-of-security/mitre-framework/mitre-attck-framework-industrial-control-systems-released/

Noman, M. (2010). Centralized and distributed anonymization for high- dimensional healthcare data. *ACM Transactions on Knowledge Discovery from Data (TKDD)* , *4*(4), 18.

O'Connell, E., Moore, D., & Newe, T. (2020). Challenges Associated with Implementing 5G in Manufacturing. *Telecom*, *1*(1), 48-67.

Radanliev, P., De Roure, D., Nurse, J. R., Nicolescu, R., Huth, M., Cannady, S., & Montalvo, R. M. (2018). Integration of cyber security frameworks, models and approaches for building design principles for the internet-of-things in industry 4.0. *Living in the Internet of Things: Cybersecurity of the IoT*, (pp. 1-6). London.

Ramsdale, A., Shiaeles, S., & Kolokotronis, N. (2020). A Comparative Analysis of Cyber-Threat Intelligence Sources, Formats and Languages. *Electronics*, *9*(824).

Riley, S. (2014, Oct 7). *"Cyber Terrain": A Model for Increased Understanding of Cyber Activity.* Retrieved from https://www.linkedin.com/pulse/20141007190806-36149934--cyber-terrain-a-model-for-increased-understanding-of-cyber-activity

SwissRE. (2019, 5). *SONAR Report: New emerging risk insights*. Retrieved from https://www.swissre.com/institute/research/sonar/sonar2019.html

Ustundag, A., & Cevikcan, E. (2017). *Industry 4.0: managing the digital transformation.* Springer.

VMware Carbon Black. (2019). *Cognitions of Cybercriminal, introducing the Cognitive Attack Loop and the three Phases of Cybersecurity.* Retrieved from https://tinyurl.com/shxr2me

Xuyun, Z. (2014). A scalable two--phase top--down specialization approach for data anonymization using mapreduce on cloud. *IEEE Transactions on Parallel and Distributed Systems* , *25*(2), 363-373.

European Cybersecurity Journal

Strategic perspectives
on cybersecurity management
and public policies

Readers' profile

- European-level representatives, sectoral agencies of the European Union, International Organisations Representatives;
- National-level officials of the Euro-Atlantic alliance, Government and Regulatory Affairs Directors & Managers;
- National and Local Government Officials as well as diplomatic representatives;
- Law Enforcement & Intelligence Officers, Military & Defence Ministries Officials;
- Legal Professionals, Representatives for Governance, Audit, Risk, Compliance, Industry leaders and innovators, active investors;
- Opinion leaders, specialised media, academic experts.

Types of contribution:

- Policy review / analysis / opinion – a Partner's article or a series of articles on crucial issues related to cybersecurity;
- Interview with Partner's representative;
- Research outcomes and recommendations;
- Advertisement of a firm, product or an event (graphical);
- Promotional materials regarding a cybersecurity conference / event (invitation, advertisement – graphical).

Do you want to share your opinion on national or European policies regarding cybersecurity? Do you want to publish outcomes of your research? Do you want to advertise?

The European Cybersecurity Journal is the right place to do it!

Prices of contribution

	PRICE (EUR)
Written contribution *Analyses, Opinions, Policy Reviews, Interviews, Research Outcomes*	100 / 1 page
Graphic contribution *Advertisement*	200 / 1 page
Graphic contribution *Advertisement*	350 / centerfold (2 pages)
Graphic contribution *Promotional campaign of an event*	250 / 1 page
Written contribution *Promotional campaign of an event*	400 / centerfold (2 pages)

CONTACT US: editor@cybersecforum.eu

The Kosciuszko Institute is a Polish think-tank founded in 2000. As an independent and non-profit organization, it gives itself the mission to contribute to the social and economic development of Poland in the European Union and as a partner of the Euro-Atlantic Alliance.

The experts of the Institute regularly cooperate with national and international organizations in the process of policy-making and initiating public debate on strategic issues.

Among its various areas of research, the Kosciuszko Institute leads its flagship project in the field of cybersecurity, within which the CYBERSEC Forum is organized.

We invite you to follow our initiatives and get involved.

THE KOSCIUSZKO INSTITUTE

is the publisher of

European
Cybersecurity
Journal