Free and Safe in Cyberspace - Aims and Backgrounder

(This document is an extensive revision of February 26th 2016 of the original Backgrounder page of the Free and Safe in Cyberspace - EU Edition 2015, published on the site on September 23rd, and distributed to the audience)

Main ambition of the *Free and Safe in Cyberspace* event series is to jointly define innovative techno-organizational **certifications and certification governance models** - within at least some current national legislative frameworks - for **ultra-high assurance IT services,** suitable for wide market adoption, as well as ***constitutional* "endpoint" lawful access systems and schemes.**

Necessarily, after Snowden and recent hacks, these new paradigms will need to assume that highly-skilled state and non-state attackers, with very limited actual liability risk, are **willing to devote tens of million of euros** to sustainably compromise its life-cycle or supply chain. They will therefore **renounce to the need or assumption of trust in anything and anyone** that is critically involved in any critical IT service life-cycle component, from certifications governance to hardware fabrication oversight; except on the assurance **quality of the overall organizational governance** of all intrinsic socio-technical constrains and (dis)incentives bearing on all entities critically involved in the entire life-cycle.

On the medium term, it is hoped that the envisioned certifications can spur substantial R&D projects and open ecosystems in a solid *actionable path* to participating actors and nations with: a renewed **digital sovereignty** of the communications of citizens and public institutions; a global **business leadership** in the most strategic security-sensitive IT sectors (such as autonomous vehicles, advanced narrow-AI, critical infrastructure, intelligence and lawful access systems); a reference for a "*trustworthy computing base*" for the **defense of critical assets and infrastructures** and strategic defense communication; a sound **low-level technological basis and governance model for *narrow* but ever wider AI systems,** in critical societal scenarios, including autonomous and semi-autonomous moveable systems.

We have identified the solution of two core challenges, Challenges A and Challenge B, as crucial to solve such crucial societal challenges and jump-start such opportunities:

CHALLENGE A: Is it feasible to provide ordinary citizens access to affordable and user-friendly end-2-end IT services with constitutionally-meaningful* levels of user-trustworthiness, as a supplement to their every-day computing devices? If so, how? What scale of investments are needed? What standards/certifications can enable a user to reliably distinguish them from other services?

CHALLENGE B: Provided that Challenge A can be met, can new voluntary international IT certifications – within some nations' current legislative frameworks – provide safeguards that are sufficiently-extreme to reconcile meaningful personal privacy, effective lawful access and prevention of malevolent use? If so, what are the core paradigms of such certification processes?

Table of Contents

1. AIMS

PRIMARY AIMS: Primary aim of the event series, in synergy with Trustless Computing Certification Campaign, is to catalyse and coordinate a wide multi-disciplinary, technically-proficient and citizen-accountable multi-stakeholder process to arrive at wide informed consensus of a critical mass of relevant actors to define and jump start **a set of new socio-technical standards, certification and certification governance bodies for critical end-2-end ICT communications, infrastructure, Artificial Intelligence, targeted lawful access systems and cyber-physical services,** that are uniquely-comprehensive and achieve unprecedented, ultra-high and *constitutionally-meaningful* levels of assurance, and assurance measurability; while preserving or increasing targeted cyber-investigation capabilities, preventing malevolent use, and overall increasing public safety.

They will enable low costs and efficient certification processes by facilitating radically open innovation models and target architectures.

In synergy with such *Campaign*, the event series aims to jump-start adequate constituent organizational processes for the future governance of such bodies, well aware that **by far the most crucial factor affecting the success in achieving and sustaining such assurance levels, is the ability to sustain extremely high-levels of technical-proficiency, citizen-accountability and presumable altruistic intentions of the key resulting decision-making bodies**. Such bodies would be international non-profits, self-financed by the costs of certifications to ICT services offered by private and public entities. They could constitute for the digital world, for example, what the *International Democratic and Electoral Assistance* represents for global elections (OMC post).

SECONDARY AIMS: Although such bodies are meant to be highly effective within current legislative and constitutional frameworks - i.e. without governmental recognition or legislative changes - they will hopefully aims to provide the **socio-technical oversight, standardization and certification basis for the enforceability** in future scenarios of recognition or adoption as voluntary or mandatory for certain classes of services by the EU or single national governments - in order to solidly comply to their Constitutions and human rights charters - by intergovernmental agreements and treaties. Examples of such treaties could the Geneva-Convention like treaty proposed by the *UN Special Rapporteur on the Right of Privacy*, the proposed *Snowden Treaty*, or standard bodies for the "so-called" *World-Sized Web* called for by Bruce Schneier. They may constitute an example (OMC post) of the "*sector-specific*" solutions to Safe Harbour issue, and other EU/US privacy issues, as suggested by Max Schrems.

Constituent processes for the creation of the mentioned intergovernmental treaties could get inspiration from those of the *Coalition for International Criminal Court*, lead by the *World Federalist Movement,* that created the International Criminal Court, or a proposed constituent process based on UN Caucuses, which was approved by the World Federalist Movement 2008 Congress (post).

2. BACKGROUNDER

2.1. CHALLENGE A - Meaningful IT Assurance

CHALLENGE A: Is it feasible to provide ordinary citizens access to affordable and user-friendly end-2-end IT services with *constitutionally-meaningful levels of user-trustworthiness, as a supplement to their every-day computing devices?** If so, how? What scale of investments are needed? What standards/certifications can enable a user to reliably distinguish them from other services?

No standards or certifications exist today that are comprehensive enough to even remotely allow a user to meaningfully assess the trustworthiness of a given end-2-end IT service. Nor any end-2-end IT service or system is available that does not contain multiple "black boxes", i.e. critical technical or organizational components which require blind trust in something or someone.

Constitutionally-meanigful?! *"While perfect assurance is impossible, we will say that a given end-2-end IT service and its related lifecycle has constitutionally-meaningful levels of trustworthiness when its levels of confidentiality, authenticity, integrity and non-repudiation are sufficiently high to make its use, in ordinary user scenarios, rationally compatible to the full and effective Internet-connected exercise of their core civil rights, except for voting in governmental elections.*

*In more concrete terms, it is end-2-end ICT service and lifecycle that warrants extremely well-placed confidence that an attacker aiming at continuous or pervasive compromisation would incur costs and risks that exceed the following: (1) for the comprimisation of the entire lifecycle, the **tens of millions of euros, significant discoverability** and unlikely liability, that are typically sustained by well-financed and advanced public and private actors, for high-value supply chains, through legal and illegal subversions of all kinds, including economic pressures; or (2) for the comprimisation of a single user, the **tens of thousands of euros,** and a significant discoverability, such as those associated with enacting such level of abuse through on-site, proximity-based user surveillance, or non-scalable remote endpoint techniques, such as NSA TAO".*

2.1.1. State of Security of commercial and high-assurance IT systems

Nowadays, IT security and privacy are a **complete debacle** from all points of view, from the ordinary citizen to the prime minister, from baby monitor to critical infrastructure, to connected cars and drones.

A **lack of sufficiently extreme and comprehensive standards** for critical computing, and the decisive covert action of states to preserve pre-Internet lawful access capabilities, have made so that, while unbreakable encryption is everywhere, nearly everything is broken; and while state-mandated or state-sanctioned backdoors are nearly everywhere[1], the most skilled or well-financed criminals communicate unchecked.

All or nearly all endpoints, both ordinary commercial systems and high-trustworthiness IT systems, are broken beyond point of encryption, and scalably exploitable by powerful nations and a relatively large number of other mid- or high-level threat actors.

EU citizens, businesses and elected state officials have no access, even at high cost, to IT and "trust services" that are NOT remotely, undetectably and cheaply compromisable by a large number of medium- and high-threat actors. **Criminal entities** that are most well-financed avoid accountability through effective use of ultra-secure IT technologies, or by relying mostly on advanced non-digital operational security techniques (OpSec). **National defenses** are increasingly vulnerable to large-scale attacks on "critical infrastructure" by state and non-state actors, increasingly capable of causing substantial human and economic harm.

The critical vulnerabilities that make nearly everything broken are nearly always either state-mandated or state-sanctioned backdoors, because the state has either created, acquired or discovered them, while keeping that knowledge hidden, legally or illegally.

Nearly all critical computing services include at least some critical components whose **complexity that is way beyond adequate verifiability**. Design or fabrication of critical components or processes (CPU, SoC fabrication, etc.) are not publicly verifiable, and there are **no reasons to trust providers' carefulness and intent**, **when plausible deniability is very easy, liability is almost non-existent, and state pressures to accidentally leave a door open are extremely high.**

Everything is broken mostly because of two structural, and highly interlinked, problems:

1. **The lack of sufficiently extreme and comprehensive standards for high-assurance IT services** that provide meaningful confidence to end-users that the entire life-cycles of their critical components are subject to oversight and auditing processes, that are comprehensive, user-accountable, publicly-assessable, and adequately intensive relative to complexity.
2. **The decisive actions taken by state security agencies to maintain pre-Internet lawful access capabilities–** since the popularization of algorithmically-unbreakable software encryption in the 1990s – through huge sustained investments in the discovery and creation of critical vulnerabilities, permeate the life-cycle and supply chain of virtually all ordinary and high-assurance IT technologies. Furthermore, the covert nature of such programs has allowed such agencies (and other advanced actors) for decades to remotely and cheaply break into virtually all end-points thought to be safe by their users – with extremely vague accountability – as well as covertly overextend their preventive surveillance capacities.

After Snowden, nearly all IT privacy activists are up in arms to fight what they see as a 2nd version of the 1990s' Crypto Wars to prevent backdoors in IT systems from being mandated by nations in the wake of "terrorism threats". Meanwhile, they overwhelmingly refrain from proposing anything about what we should do about those state backdoors and critical vulnerabilities that already exist everywhere. Almost no-one challenges state security agencies pretence that they are "going dark" to trumpet how much they are missing capabilities to enable lawful access, when they overwhelmingly are not, not even for scalable targeted attacks. Most activists are focused on:

- (a) pushing existing Free/Open Source Software privacy tools to the masses, while making them more user-friendly and incrementally safer with small grants (no matter if from USG) and

- (b) going full-out there to fight a 2nd Crypto War to prevent the government to create a state backdoor.

First off, the **Crypto Wars in the 1990's were won in appearance, but utterly lost in essence**. In fact, while the US and other governments backtracked on their proposal for an ill-conceived mandatory backdoor (such as the Clipper Chip) and algorithmically-unbreakable encryption became accessible to anyone, the most powerful states security agencies won many times over. In fact, over the following two decades:

- (1) powerful state security agencies have surreptitiously and undetectably placed backdoors nearly everywhere, with nonexistent or very insufficient due process oversight, compared to the already inadequate oversight lawful interception systems;
- (2) Tons of valuable targets, even very "up there", have kept using IT devices that they thought lacked backdoors, but which had been snooped upon for years or decades without their knowledge;
- (3) The general perception that "free crypto for all" had won prevented even a demand for meaningful IT devices to be developed, which could minimize, isolate, simplify and do away with the need of trust in very many of untrustworthy actors along critical phases of the device life-cycle.

A mix of government self-serving disinformation and self-interested over-representation of the strength of current FLOSS solutions has brought many to believe that endpoint exploitation of well-configured FLOSS device setup is not scalably exploitable by advanced attackers.

Many of such IT privacy activists and experts over-rely on NSA documents' reference to the fact that some advanced attackers take care not to overuse certain exploitation techniques to avoid burning them. But careful analysis of the capabilities of NSA Turbine, NSA FoxAcid, Hacking Team RCS document shows how advanced endpoint exploitation techniques allow them to scale highly and prevent such "burning" risk by using exploits that are beyond the ability of the target to discover, and other techniques. More recently, one leader of such activists, Jacob Appelbaum, said clearly (min 37.15-38.15): "*It does seem to indicate to me that cryptography does stop them ... I have seen that the Tor browser stops them from doing passive monitoring, and they have to switch to targeted. And that's good. We want them to go from bulk, or mass, surveillance to targeted stuff. Now, the targeted stuff, because it is automated, is not different in scale but just different in methodology, .. actually. And usually they work together*"

All the while, the media success of such security agencies in wildly overstating the "going dark" problem has enabled them to gather substantial political and public opinion consensus for:

- (1) unconstitutional surveillance practices gravely affecting non-suspect citizens, and often setting up multiple redundant legal authorities;
- (2) convincing politicians and the public opinion of the need to "outlaw" encryption and/or extend inadequate lawful access mandate, traditionally reserved to telephone operators, to all digital communications.

All the while, by breaking everything, they expose US government, US private interests, and law enforcement and intelligence to grave damage for state security and espionage by foreign states and non-state actors.

Impact on IT Security Business

EU IT security/privacy service/systems providers are increasingly unable to sustainably compete and innovate as they are **unable to differentiate on the basis of meaningful and**

comprehensive benchmarks. They are also increasingly unable to convince users to investing in fixing vulnerabilities in one part of their systems, when most-surely many vulnerabilities remain in other critical parts, which are known to the same kind of threat actors. In a post-Snowden world, the success of even high-assurance cyber-security systems is increasingly "**security theatre**", because even the highest-assurance systems in the civilian market contain at least one critical vulnerability, accessible in a scalable way by even mid-level threat actors, with very low risk of discoverability and attribution.

So therefore it is almost impossible to measure and sustain the current overall security added value of any new security service, and related risk management strategies, even before assessing the increase in attack surface and vulnerabilities that any new product entails. The only reliable measure of the effectiveness of an high-assurance IT security provider, private and public, relies on its "closeness" to major stockpilers of vulnerabilities, mostly few large powerful states, creating pervasive intelligence network effects[2], gravely undermining society sovereignty, freedoms and competitiveness.

As blogger Quinn Norton said, everything is broken. Revelations about systems and programs like NSA Turbine, NSA'S FoxAcid and Hacking Team, have shown the huge scalability - in terms of low risk and cost - of completely compromising of endpoint devices, by numerous public and private actors, and even more numerous actors that do or could trade, lend or steal such capabilities. It's become clear that no IT system that assumes need for trust in any one person or organization - and there are none today - can be considered meaningfully trustworthy.

The exception to this rule is that there are some people in the world - **top criminals, billionaires, or highest state official - who do have access to devices that are most likely not compromised by external entities**. This results in a huge asymmetry of power between them and all others, i.e. two sets of citizens.

This situation will not be changed by a nation's law, or international treaty. Stockpiling of zero day vulnerabilities, through investment in discovery, creation and purchase by powerful state and non-state actors will keep accelerating, and **there is no chance any law or international treaty can significantly slow an acceleration of stockpiling of 0-days**. Non proliferation of IT weapons is very different from other weapons like biological and nuclear, as their nature makes them easier to hide and reproduce, and they are used and spread daily by powerful actors to pursue their cyber-investigation goals.

In summary, there is a wide unavailability, for both citizens and for *lawful access* schemes, of end-2-end *IT services of meaningfully high-trustworthiness levels.*

2.1.2. The problem with current certifications and certifications governance

Over the last decades, in addition to sanctioning backdoors everywhere, states have repeatedly proven to be utterly incapable of either socio-technically designing, or legally managing, or issuing proper technical and organizational certification requirements for lawful access compliance. Fittingly, states have been similarly unable to create voluntary or mandatory IT security standards that would be anything like sufficiently extreme and comprehensive.

A recent ENISA report, highlights: "*At the time of writing, there is no single, continuous 'line of standards' related to cyber security, but rather a number of discrete areas which are the subject of standardisation*". The current EU Cybersecurity Strategy (2013) calls for new standards and certification schemes - including supply chains and hardware - and calling for "*a renewed emphasis on dialogue with third countries, with a special focus on like-minded partners that*

share EU values", "*possibly establish voluntary EU-wide certification schemes building on existing schemes in the EU and internationally*". Recognizes "*the need for requirements for transparency, accountability and security is becoming more and more prominent*". Nonetheless, ECSEL and EDA no "EU computing base" exist in most IT domains that is even publicly verifiable in all its critical parts.

Consensus-based decision making processes at the core of EU institutions – and international public and mixed standard-setting bodies (such as ETSI) – have made it impossible to resist the firm will of even a single powerful country to corrupt or dilute-to-meaninglessness the standard setting process.

Industry driven standards are no better; standards bodies like *Trusted Computing Group* and *Global Platform* have focused on increasing user convenience, interoperability and reducing overall costs of violations to content copyright and integrity of financial transactions, while playing lip service to the security and privacy demands of end-users that were at odds with state security agencies.

So therefore, the governance of such paradigms and certifications may need to be primarily independent, international, highly-competent and citizen-accountable, and the role of the national, and international governmental institutions (EU, UN, etc) – and major global IT industry players – can only be that recognizers, adopters, and minority stakeholders. A process similar to that of the World Wide Web Consortium could be followed, but with much wider user- or citizen-accountability to keep companies from having too much control.

2.2. CHALLENGE B - Constitutional Lawful Access

Provided that Challenge A can be met, can new voluntary international IT certifications - within some nations' current legislative frameworks - provide safeguards that are sufficiently-extreme to reconcile meaningful personal privacy, effective lawful access and prevention of malevolent use? If so, what are the core paradigms of such certification processes?

Related questions:

- What are the effects on public safety and wellbeing of the current unavailability of IT devices that are reliably resistant to undetected remote compromisation by mid- or high-threat actors, legal or illegal?

- What are the foreseeable effects on public safety and wellbeing of the wide availability of such IT devices, therefore resistant even to remote access by public security agencies with legal due process?

- Could new lawful access schemes for high-assurance IT services and systems rely, not on states, but on provider-managed voluntary "key recovery" schemes certified by “trustworthy 3rd parties”, such as radically citizen-accountable, independent and competent international bodies?

- Could the inevitable added risk be essentially shifted from technical systems to on-site organizational processes?

World citizens are continuously asked by states and by digital rights activists: "Would you rather be free or safe?". **We suspect digital privacy and public safety are not an 'either or' question, but instead a 'both or neither' challenge.** Albeit acknowledging that solving one or both of the challenges may not be possible, we believe extensive resources intellectual and monetary should be devoted to such attempt. Refusing to make counterproposals that acknowledge the crucial need for constitutional lawful access and attempt to take it into account - as most digital rights activist organizations have done to date - is backfiring for digital civil rights by ensuring that governments go ahead maintaining the status quo or implement"sub-optimal" standards and laws.

EU Cybersecurity Strategy (2013) calls for “*The same laws and norms that apply in other areas of our day-to-day lives apply also in the cyber domain.*”. “*Fundamental rights, democracy and the rule of law need to be protected in cyberspace. But freedom online requires safety and security too*”

2.2.1. Assurance of current lawful access schemes

Today, state-mandated backdoors – hidden or public like the telephone interception systems – or *state-sanctioned backdoors* - such as undisclosed critical vulnerabilities created, acquired, discovered or used, legally or illegally – are in nearly all IT devices.

Over the last decades, in addition to sanctioning backdoors everywhere, individual states have repeatedly proven to be utterly incapable of either socio-technically designing, or legally managing, or setting proper technical and organizational requirements for state lawful access compliance. Nonetheless, the dire need to reconcile privacy and cyber-investigation remains as crucial as ever.

US and most western states have in place plenty of legislations, and legally authorized intelligence programs, that enable them to access a suspect communications following a legal due process authorization, including: mandatory key disclosure, lawful hacking laws, national security letters, and other laws.

Powerful states invest tens of millions of dollars every year in pressure of all kinds order to ensure that *IT systems of meaningfully high-trustworthiness levels* are not available to the civilian market and, indirectly, to nearly all of the internal intelligence, military and lawful access systems markets. Such pressures are in the form of creation and discovery of symmetrical backdoor, onsite subversion of various kinds, economic (CIA venture capital, procurement pressures, etc), patenting (NSA secret patents), legal (crypto export) pressure crypto export pressures, and strong pressures to establish high-trustworthiness IT standards, that are incomplete (Common Criteria, FIPS, etc.) and compromised (Dual_EC_DRBG). That is in addition to similar activities by other powerful states, and tens of millions of euros of investments by zero market companies.

Nonetheless, a few of the most knowledgeable and well-funded criminals, state and non-state, regularly do use and could use custom-made end-2-end IT infrastructures that manage to avoid the use of components where critical vulnerabilities are known by powerful states. On the other hand, commercial vendors like Apple - having uniquely full control of their life-cycle, and not being mandated to store a master key - are in theory positioned to render their future systems inaccessible to lawful access. However that is very unlikely because of: the huge relative complexity of their systems and life-cycle, which makes it inherently creation of weakness via subversion, legal or illegal by powerful state actors, as well as to independent discovery of vulnerability; and high-level of plausible deniability in a scenario in which Apple may be purposely leaving highly-safeguarded and asymmetrical backdoors for a few states. The same arguments are valid for current high-assurance IT systems, which in all known case add the lack of control of a number of critical life-cycle phases.

2.2.2. Current best practices, approaches and proposals

Current best practices in certain states

The Brazilian state IT agency SERPRO has internal regulations that intrinsically requires 4 state officials of 4 different public agencies need to be physically present at a specific hosting room and consent in order to allow access to the emails of a state employee based on a court order. More recently, they are increasing the assurance of their solution for both citizens and law enforcement on the server side with additional safeguards, through Kryptus solutions. Such an approach, however, still does not deal adequately with the assurance of several other potential vulnerabilities in the life-cycle, such as: client devices HW-SW, other critical SW and HW stack on the server side, the systems use by law enforcement to manipulate and store the acquired info, hardware fabrication of critical HW components.

The law enforcement access to a user's keys in Austria for digital passports currently require 3 officials from different state agencies in in-person secret-sharing and "threshold secret" processes.

In addition to much higher and more comprehensive assurance requirements - given the law trust in government - citizen-witnesses or citizen-juries may want to be added to the officials from different state agencies, in order to add an additional layer of guarantee!

Dominant strategy of mainstream digital civil rights activists

Almost all citizens and many activists recognize the benefits of enabling due process lawful access for criminal investigation, but the grave incompetence and abuse by states have brought most experts to believe that such access cannot be ensured without unacceptable risks for citizens' liberty.

The IT security industry is creating solutions that either are based on or add to systems which are non verifiable in critical parts, and whose complexity is way beyond what can ever be sufficiently audited. Meanwhile, IT privacy activists push similarly inadequate existing Free and Open source privacy tools to the masses, while just increasing usability, or at best seeking inadequate small grants for very inadequate complexity reduction, and increases in isolation and auditing.

Proposals "hinted at" by US/UK governments

In recent statements, NSA, Europol, UK Cameron, Obama, US Dept of Justice, and FBI have proposed to solve the "going dark" problem by mandating a some kind of backdoor into all IT systems. The FBI has more specifically proposed a "*legislation that will assure that when we get the appropriate court order . . . companies . . . served . . . have the capability and the capacity to respond*", while the NSA has been generically referring to organization or technical safeguards ensuring backdoor access authorization approval by multiple state agencies[5], and Obama referring to a possible safeguard role of non-state entities[6].

From Snowden and Hacking Team revelations, it has become clear that - in addition to covertly introducing, purchasing and sanctioning symmetric backdoors everywhere - most western nations have consistently proven incapable or unwilling to design, standardize, legally oversee or certifying *lawful access*, by LEA or intelligence agencies, both for traditional phone wiretaps and for IT systems. Current schemes and systems have very poor or no citizen or legislative-branch accountability, because of lack of legal mechanisms as well as adequately accountable socio-technical systems.

Such precedents and a number of technical facts make so that such solution would most likely turn out to be ineffective towards the most serious criminals and causing great risks for civil liberties abuse[7]. Among the infeasibilities is the fact that - short of mandating a complete and impossibly draconian control over any connected IT devices through unbreakable remote attestation - how can any master-key for *lawful access* in IT products prevent a suspect to encrypt its messages a second time, possibly through steganography, rendering the master-key useless in reading the plain text or audio, and even hard to prove the suspect has sent an unlawfully encrypted message?

Proposals for recommendations by EU Parliament

A new report has been commissioned by the EU Parliament which seems to advice that "lawful cracking" lawful access systems **"when they are used in Europe with the appropriate oversight and safeguards could have legitimate purposes"**. Although, currently such systems appear to unconstitutional in Italy and Germany (except for intelligence purposes) for example, though they are legal in the US.

"No new state-mandated backdoor" proposal, by 14 leading US/UK IT security experts (1997-2014)

In an open letter published last July 6th 2015, Keys under Doormats - 14 among the most renowned US computer security experts have made a detailed case against the introduction of new national legislations in the US and elsewhere, and possibly part of international agreements. They also list questions that any such proposal should answer in order for the public and experts to assess the foreseeable risks of grave civil liberties abuses. The document follows is intended as an upgrade to recent development of a very extensive and influential similar proposal from the 1990's, The Risks of Key Recovery, Key Escrow, and Trusted Third-Party Encryption.

Proposal for formalization and regulation of "lawful cracking" by Bellovin, Maze, Landau and Clark

Even some IT security experts that have been for decades the most staunch opposers to lawful access solution for IP communications, acknowledge that some "going dark" problem exists or could potentially exist and - regardless of quite varying opinions about its gravity - a solution will need to be found as political pressures will keep mounting[8].

Three of the most prominent among the 14 experts mentioned above, and Sandy Clark, have proposed [through Going Bright, 2013, and *Lawful Hacking: Using Existing Vulnerabilities for Wiretapping on the Internet*, 2014] an alternative solution to the problem that requires the state to *"exploit the rich supply of security vulnerabilities already existing in virtually every operating system and application to obtain access to communications of the targets of wiretap orders",* and properly regulate it. It basically proposes to formalize and strictly regulates the state's ability to hack citizens pursuant a court order. It proposes very extensive measures and safeguards to mitigate the consequent negative effects, including:

1. Creation of new vulnerabilities is not allowed, but only discovery and creation of exploit for existing vulnerabilities.
2. Mandatory reporting of vulnerabilities to IT vendors on discovery or acquisition, with some exceptions. It counts on the fact that new will be found and that it takes time for vulnerabilities to be patched;
3. Limitation of lawful access software to only authorized access actions (whether intercept, search, or else).

They propose to formalize and regulate the use of "lawful cracking" techniques as a way to enable the state to pursue cyber-investigation:

"We propose an alternative to the FBI's proposal: Instead of building wiretapping capabilities into communications infrastructure and applications, government wiretappers can behave like the bad guys. That is, they can exploit the rich supply of security vulnerabilities already existing in virtually every operating system and application to obtain access to communications of the targets of wiretap orders.

We are not advocating the creation of new security holes, but rather observing that exploiting those that already exist represents a viable—and significantly better—alternative to the FBI's proposals for mandating infrastructure insecurity. Put simply, the choice is between formalizing (and thereby constraining) the ability of law enforcement to occasionally use existing security vulnerabilities—something the FBI and other law enforcement agencies already do when necessary without much public or legal scrutiny or living with those vulnerabilities and intentionally and systematically creating a set of predictable new vulnerabilities that despite best efforts will be exploitable by everyone."

2.3. QUESTION A: What is the role of the Free Software in ultra-high assurance IT standards?

What is the role of the Free/Open Source Software movement for the prospects of wide availability of computing with meaningful user control.

Over the last thirty years, a huge amount of volunteer and paid work has been devoted to developing Free Software with the aim of promoting users' civil freedom in computing.

Why then, to date, is no end-user computing device available at any cost which would give the user meaningful confidence that its computing is not completely compromised undetectably at insignificant cost and risk?

Why is no end-user device available today that does NOT contain at least some "critical" software/firmware components that (a) are not nearly sufficiently verified relative to complexity? or (b) are non-verifiable in its source code (without NDA) or even proprietary?

What should be the free software community priorities and short and long-term objectives in a Post-Snowden World?

Free/Open Source Software, while providing important civil freedom, does not *directly* improve trustworthiness of a software application or stack, in comparison to that whose source code is merely publicly-verifiable without NDA. At times, on the contrary, it has constrained available business models in ways that prevented the sustainable attraction of the very large resources necessary to guarantee a sufficiently-extreme auditing relative to complexity.

Nonetheless, an adequate new standard may need to **very strictly mandate Free/Open Source Software and firmware, with few and/or temporary exceptions for non-critical parts**, because it strongly promotes incentives for open innovation communities, volunteer expert auditing and overall ecosystem governance decentralisation.

These, in turn, substantially contribute to IT actual and perceived security, and promotes an **ecosystem that is highly-resilient to very strong economic pressures**, as well as short- and long-term changing technological, legislative and societal contexts.

Most importantly, without the very active and well meaning participation (paid and not paid) of many of the world-best IT security experts and “communities”, it would be unlikely to achieve a sufficiently-extreme **necessary auditing intensity and quality, relative to complexity** that is needed to achieve the project aims. Without such participation, it would be unlikely that a project even with a budget of over hundreds of millions of euros could have reasonable expectations to prevent successful remote attacks from the numerous and varied entities, which have access to remote vulnerabilities that are regularly devised, commissioned, acquired, purchased or discovered, by entities that are extremely well-financed, have unprecedented accumulated skill-sets and often low or inexistent actual liability.

2.4. QUESTION B: Can ultra- high-assurance IT certifications and governance models help promote short and long-term AI safety?

Can standards for radically more trustworthy IT define a European *actionable path*, from the short to the long-term, to: (1) restore meaningful digital sovereignty to EU citizens, businesses and institutions; (2) cement a EU leadership in the most security-sensitive IT and Artificial Intelligence sectors (such as autonomous vehicles, surveillance, etc.); and (3) substantially increase the chances of utopian rather than dystopian long-term artificial intelligence prospects?

In recent years, rapid developments in AI specific components and applications, theoretical research advances, high-profile acquisitions from hegemonic global IT giants, and heart-felt declarations about the dangers of future AI advances from leading global scientists and entrepreneurs, have brought AI to the fore as both (A) the **key to private and public economic dominance in IT, and other sectors**, in the short-to-medium term, as well as (B) the leading **long-term existential risk (and opportunity) for humanity**, due to the likely-inevitable "machine-intelligence explosion" once an AI project will reach human-level general intelligence.

Google, in its largest EU acquisition this year acquired for 400M€ a global AI leader, DeepMind; already invested by Facebook primary initial investors Peter Thiel and by Elon Musk. **Private investment in AI has been increasing 62% a year**, while the level of secret investments by multiple secretive agencies of powerful nations, such as the NSA, is not known – but presumably very large and fast increasing – in **a winner-take-all race to machine super-intelligence among public and private actors, which may well already have started**.

A recent survey of AI experts estimates that there is a 50% chance of achieving human-level general intelligence by 2040 or 2050, while not excluding significant possibilities that it could be reached sooner. Such estimates may even be biased towards later dates because: (A) there is an intrinsic interest in those that are by far the largest investors in AI – global IT giants and USG – to avoid risking a major public opinion backlash on AI that could curtail their grand solo plans; (B) it plausible or even probable that substantial advancements in AI capabilities and programs may have already happened but have successfully kept hidden for many years and decades, even while involving large numbers of people; as it has happened for surveillance programs and technologies of NSA and Five Eyes countries.

Some of the world's most recognized scientists, and most successful technological entrepreneurs believe that progress beyond such point may become extremely rapid, in a sort of "intelligence explosion", posing grave questions on humans' ability to control it at all. (See Nick Bostrom TED presentation). Very clear and repeated statements by Stephen Hawking (the most famous scientist alive), by Bill Gates, by Elon Musk (main global icon of enlightened tech entrepreneurship), by Steve Wozniak (co-founder of Apple), agree on the exceptionally grave risks posed by uncontrolled machine super-intelligence.

Elon Musk, shortly after having invested in DeepMind, even declared, in an erased but not retracted comment:

"*The pace of progress in artificial intelligence (I'm not referring to narrow AI) is incredibly fast. Unless you have direct exposure to groups like Deepmind, you have no idea how fast-it is growing at a pace close to exponential. The risk of something seriously dangerous happening is in the five-year timeframe. 10 years at most. This is not a case of crying wolf about something I don't understand.I am not alone in thinking we should be worried. The leading AI companies have taken great steps to ensure safety. They recognise the danger, but believe that they can shape and control the digital super-intelligences and prevent bad ones from escaping into the Internet. That remains to be seen…*"

Some may argue why extreme IT security to support AI safety is needed now if its consequences may be far away. One clear and imminent danger is posed by self-driving and autonomous vehicles (aerial and terrestrial) - which utilize increasingly wider narrow AI systems - and **the ease with which they can be "weaponized" at scale**. Hijacking the control of a large number of drones or vehicles could potentially cause hundreds of death or more, or cause hardly attributable hacks that can cause grave unjustified military confrontation escalations.

Richard Hawkings summarised it most clearly when he said **"Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all".** Control relies on with IT assurance to ensure that who formally controls it is also who actually controls it, and possibly with international IT assurance certification governance that may provide a governance model for international efforts to regulate advanced AI projects or better to guide international democracy projects to develop "friendly AI" before unfriendly AI gets to human-level general intelligence.

On Jan 23rd 2015, nearly the entire "who's who" of artificial intelligence, including the leading researchers, research centers, companies, IT entrepreneurs – in addition to what are possibly the leading world scientists and IT entrepreneurs – signed Open Letter "Research priorities for robust and beneficial artificial intelligence" with an attached detailed paper.

Such an Open Letter, although it is a greatly welcome and needed general document, overestimates the levels of trustworthiness and comparability of existing and planned high-assurance IT standards, as well as the at-scale costs of "high-enough for AI" assurance levels, and focuses on security research as a way to make AI "more robust".

Such an Open Letter emphasizes the need for "more robust AI". A **very insufficiently-secure AI system may be greatly "robust" in the sense of business continuity, business risk management and resilience**, but still be extremely weak in safety or reliability of control. This outcome may sometimes be aligned with the AI sponsor/owner goals – and those of other third parties such as state security agencies, publicly or covertly involved – but be gravely misaligned to chances to maintain a meaningful democratic and transparent control, i.e. having transparent reliability about what the system, in actuality, is set out to do and who, in actuality, controls it.

More important than "robustness", sufficiently-extreme security assurance levels may comprise the most crucial foundation for AI safety in the short and long terms, and serve to increase **transparency of who is *actually* in control**, as well as a precondition for verification and validity.

As AI systems are used in an increasing number of critical roles, they will take up an increasing proportion of cyber-attack surface area. It is also virtually certain that AI and machine learning techniques will themselves be used in cyber-attacks. There is a large amount of evidence that many advanced AI techniques have long been and are currently being used by the most powerful states intelligence agencies, to attack – often in contrast with national or international norms – end-users and IT systems, including IT systems using AI. As said above - while the levels of **investment of public agencies of powerful nations, such as the NSA, is not known - it is presumably very large and fast increasing**, in a possibly already started race among public and private actors. A race that could in the near future accelerate into a winner-take-all race. The distribution of such funding by secretive state agencies, among offensive R&D rather than defensive R&D (i.e. AI safety), will most likely follow the current ratio of tens of times more resources for offensive R&D.

The above Open letter states that *"Robustness against exploitation at the low-level is closely tied to verifiability and freedom from bugs"*. This is correct; but may be only partially so, especially for use in critical and ultra-critical use cases, which will become more and more dominant. It is better to talk about auditability in order not get confused with (formal) IT verification. It is crucial and unavoidable to have complete auditability and extremely diverse, competent and well-meaning actual auditing of all critical HW, SW and procedural components involved in an AI system's life-cycle, from certification standards setting, to CPU design, to fabrication oversight. (Such auditing may need to happen in secrecy, because public auditability of SW and HWs design may pose a problem in so far as project pursuing "unfriendly Super-intelligence" could get advantage in a winner-take-all race). Since 2005 the US Defense Science Board has highlighted how "Trust cannot be added to integrated circuits after fabrication" as vulnerabilities introduced during fabrication can be impossible to verify afterwards. Bruce Schneier, Steve Blank, and Adi Shamir, among others, have clearly said there is no reason to trust CPUs and SoCs (design and fabrication phases). No end-2-end IT system or standards exist today that provide such complete auditability of critical components.

It is **impossible, and most probably will remain so, to ensure perfectly against critical vulnerabilities** - given the socio-technical complexity of IT socio-technical systems - even if they were to be simplified by 10 or 100 times, and with radically higher levels of auditing relative to complexity.

Nonetheless, it remains extremely crucial and fundamental that adequate research **could devise ways to achieve sufficiently-extreme level confidence about "freedom from critical vulnerabilities" through new paradigms.** We may need to achieve sufficient user-trustworthiness that sufficient intensity and competency of engineering and "auditing efforts relative to complexity" have been applied, for all critical software and hardware component. **No system or standard exist today to systematically and comparatively assess** – for such target levels of assurance for a given end-2-end computing service, and its related life-cycle and supply-chain.

As stated above, all AI systems in critical use cases – and even more crucially those in advanced AI system that will soon be increasingly approaching machine super-intelligence – will need to be so robust in terms of security to such an extent that they are **resistant against multiple extremely-skilled attackers willing to devote cumulatively even tens or hundreds of millions of Euros to compromise at least one critical components of the supply chain or lifecycle**, through legal and illegal subversion of all kinds, including economic

pressures; while having high-level of plausible deniability, low risk of attribution, and (for some state actors) minimal risk of legal consequences if caught.

In order to **reduce substantially these enormous pressures**, it may be very useful to research socio-technical paradigms by which sufficiently-extreme level of AI systems user-trustworthiness can be achieved, **while at the same time transparently enabling due legal process cyber-investigation and crime prevention**. The possible solution of such dichotomy would reduce the level of pressure by states to subvert secure high-assurance IT systems in general, and possibly – through mandatory or voluntary standards international lawful access standards – improve the ability of humanity to conduct cyber-investigations on the most advanced private and public AI R&D programs. **Cyber-investigation may be crucial to investigate some criminal activities that aimed at jeopardizing AI safety efforts.**

There is a need to **avoid the risk of relying for guidance on high-assurance low-level systems standard/platform projects from defense agencies of powerful nations**, such as the mentioned DARPA SAFE, NIST, NSA Trusted Foundry Program, DARPA Trust in Integrated Circuits Program, when it is widely proven that their intelligence agencies (such as NSA) have gone to great lengths to **surreptitiously corrupt technologies and standards**, even those that are overwhelmingly used internally in relatively high-assurance scenarios.

The **cost of radically more trustworthy low-level systems for AI could be made very comparable to commercial systems** mostly used as standard in AI systems development. Those cost differentials could possibly be reduced to being insignificant through production at scale, and open innovation models to drive down royalty costs. For example, hardware parallelization of secure systems and lower unit costs (due to lower royalties), could make so that adequately secure systems could compete or even out compete in cost and performance those other generic systems.

There is a lot of evidence to show that R&D investment on solutions to defend devices from the inside (that assume inevitable failure in intrusion prevention) could end up increasing the attack surface if those systems life-cycle are not themselves subject to the same extreme security standards as the low level system on which they rely. Much like antivirus tools, password storing application and other security tools are often used as ways to get directly to a user or end-point most crucial data. The recent scandals of NSA, Hacking Team, JPMorgan show the ability of hackers to move inside extremely crucial system without being detected, possibly for years. DARPA high-assurance program highlight how about 30% of vulnerabilities in high-assurance systems are introduced by internally security products.

Ultimately, it may be argued that IT assurance that is high enough for critical scenarios like advanced AI is about the competency and citizen-accountability of the organizational processes critically involved in the entire life-cycle, and the intrinsic constraints and incentives bearing on critically-involved individuals within such organizations.

Maybe, the dire short-term societal need and market demand for radically more trustworthy IT systems for citizens privacy and security and societal critical assets protection can be aligned in a grand strategic vision for EU cyberspace to satisfy – in the medium and long-term – both the huge societal need and great economic opportunity of creating large-scale ecosystems able to produce AI systems that will be high-performing, low-cost and still provide adequately-extreme levels of security for AI most critical usage scenarios.

It is worth considering if short-term and long-term R&D needs of artificial intelligence "("AI") and information technology ("IT") – in terms of security for all critical scenarios – may become

synergic elements of a common “short to long term” vision, producing huge societal benefits and shared business opportunities. The dire short-term societal need and market demand for radically more trustworthy IT systems for citizens’ privacy and security and societal critical assets protection, can very much align – in a grand strategic cyberspace EU vision for AI and IT – with the medium-term market demand and societal need of large-scale ecosystems capable to produce AI systems that will be high-performing, low-cost and still provide adequately-extreme levels of security for AI critical scenarios.

2.5. TOWARDS A SOLUTION for Challenges A and B

2.5.1. Are wide investments in Challenge A realistic or sustainable in the absence of a concurrent solution to Challenge B?

It has emerged that almost all western nations, including the US and most EU countries, have one or more lawful ways under which state security agencies can intrude on the privacy of millions of citizens without a court order, including some sort of mandatory key disclosure legislation.

Although their existence has been hidden, they are currently politically based on two justifications:

- (1) to preserve or restore the traditional lawful intercept capabilities that have been lost in cyberspace;
- (2) to perform dragnet or large-scale targeted surveillance in order to prevent or prosecute grave crimes.

The surprising public acceptance of the second justification is arguably dependent on the fact that currently it is the only way to achieve the first justification. The auspicable wide-market uptake of new high-assurance ICT standards and solutions is a challenge that may need to be solved concurrently with the challenge of devising ways to restore legitimate criminal investigation capabilities in cyberspace. It has become clear that citizens will choose perceived physical safety over cyber-privacy if given a stark choice. In fact, political and public opinion pressures to extend the outlawing or allowing subversion of such techs and standards would be huge, as they've been for decades, and would surely increase to become unbearable after major terrorist attacks, largely attributed to the use of such technologies. Such grave risks of legal sustainability comprise a major obstacle to private investment and public commitments in wide deployment of such high-assurance technologies and standards for the civilian market.

2.5.3. Current legislative frameworks

The legislation affecting high-assurance IT systems and lawful access systems is constituted of national laws, which are mildly influenced in their regulatory implementations by voluntary international public and/or private standards (Common Criteria, ETSI, NIST, etc.).

High-assurance IT services are regulated with the over-riding aim to prevent malevolent use, and therefore focused on limiting export (crypto export laws) and use of certain technologies, and increasingly their research (such as in the ongoing Wassenaar Agreement national implementations).

Lawful access processes, in both state security and civilian scenarios, are instead subject to very limited or inexistent technical regulation of the security of their technical infrastructure against abuse by state agencies on their citizens, or by external actors against such state agencies. They are subject to articulated, though largely inadequate, organizational and socio-technical regulations and oversight procedures. Multiple "Mutual legal assistance" treaties regulate the often-crucial international cooperation in pursuit of cyber-crimes. Such arrangements are so insecure and slow, taking often months, that most times workarounds "at the edge of legal" are deployed[3] , which are overwhelmingly unregulated.

2.5.4. Linking Challenges A and Challenge B

Although solving Challenge A would provide substantial societal benefits, the concurrent solution of Challenge A and Challenge B would provide substantially higher and more-sustainable societal benefits because of: the interdependency of constitutional mandates for public safety and constitutional rights for personal privacy; the need to reduce the chance of abuse of Challenge A by criminals, including third state actors; and the need for effective provisioning and investments in Challenge A to be legally sustainable, even though grave public safety crimes might be substantially aided, allegedly or actually, by the use of Challenge A.

Most law enforcement agencies (LEAs) in their (possibly self-serving) public claims and most e-privacy experts believe that Challenge A is already available to ordinary citizens, willing to sacrifice substantial money and/or usability. Almost all privacy experts believe Challenge B is completely impossible, and all discussions or proposals to find a solution are either nonsense, insincere or both.

Many believe that Challenge A is impossible or very uneconomical.

Most privacy experts, even those privately admitting there may be some way to solve Challenge B right, believe that such a possibility is so remote that we should not publicly investigate it, as it would increase the risk that states may deploy the wrong solutions.

A few experts believe that Challenge A is possible, or economical, but will never be or should never be sustainably widely available unless Challenge B is also substantially solved. Almost all LEAs and very few privacy experts believe Challenge B may be feasible by deeply exploring innovative socio-technical paradigms, relying on concepts such as: secret-sharing, multi-party computation in different jurisdictions, secret-sharing relying on-site processes rather than IT, provider management, independent standardization and oversight, citizen-witness processes, and more.

2.5.5. Related questions

- Could the solution to such challenges – through the creation of international certification processes and open and resilient ecosystems – cement a future global leadership of EU values and EU industry in the most security- and privacy-sensitive areas of IT, such as personal communications, state security and defense, IoT and advanced artificial intelligence? How much of the new paradigms needed to solve Challenge A can help solve Challenge B?
- Can solving Challenge A be legally sustainable unless we solve Challenge B? Can the wide-scale investments needed to bring meaningful privacy to all be secured if these are legally unsustainable in time?
- Can the feasibility of solving such problems be dramatically reduced by aiming at computing services that are supplementary to current ordinary commercial devices? A sort of meaningfully private sphere, though feature-wise limited, alongside a digital public sphere for more general computing?
- What are the effects on public safety of the current wide unavailability of meaningfully secure IT devices? What are the effects on public safety and public interest of a possible future wide availability of meaningfully secure IT, and therefore resistant to scalable remote access by even public security agencies through lawful access requests?
- Can independent citizen-accountable, citizen-witness or citizen-jury organizational processes - from standard setting, to fabrication and key recovery oversight - substantially

or radically increase the actual and perceived trustworthiness of setting standards and critical lifecycle phases of IT devices?

- Can a critical mass of international actors lead the creation of independent, citizen-accountable new standard, platforms and ecosystems for trustworthy IT that can underpin EU values and EU business global leaderships in security- and privacy-critical computing?
- Can an actionable path for Europe envision, from the short to the long term, to radically restore the access by citizen and businesses to private civic communications, to safeguarding critical defense infrastructure, to provide a unique competitive advantage, and long-term safety, for the most future critical EU Artificial Intelligence projects and services?

Authors

Rufo Guerreschi

Exec. Dir. Open Media Cluster

Jovan Golic

EIT Digital Action Line Leader for Privacy, Security and Trust

Telecom Italia Information Technology