Science and Engineering Ethics (2020) 26:1267–1285
https://doi.org/10.1007/s11948-019-00138-8

ORIGINAL RESEARCH/SCHOLARSHIP

Ethical Value-Centric Cybersecurity: A Methodology Based on a Value Graph

Josep Domingo-Ferrer[1] · Alberto Blanco-Justicia[1]

Received: 17 January 2019 / Accepted: 16 September 2019 / Published online: 30 September 2019

Abstract
Our society is being shaped in a non-negligible way by the technological advances of recent years, especially in information and communications technologies (ICTs). The pervasiveness and democratization of ICTs have allowed people from all backgrounds to access and use them, which has resulted in new information-based assets. At the same time, this phenomenon has brought a new class of problems, in the form of activists, criminals and state actors that target the new assets to achieve their goals, legitimate or not. Cybersecurity includes the research, tools and techniques to protect information assets. However, some cybersecurity measures may clash with the ethical values of citizens. We analyze the synergies and tensions between some of these values, namely security, privacy, fairness and autonomy. From this analysis, we derive a value graph, and then we set out to identify those paths in the graph that lead to satisfying all four aforementioned values in the cybersecurity setting, by taking advantage of their synergies and avoiding their tensions. We illustrate our conceptual discussion with examples of enabling technologies. We also sketch how our methodology can be generalized to any setting where several potentially conflicting values have to be satisfied.

Keywords Cybersecurity · Ethics · Privacy · Fairness · Autonomy

The following funding sources are gratefully acknowledged: European Commission (Project H2020 700540 “CANVAS”); Government of Catalonia (ICREA Acadèmia Prize to J. Domingo-Ferrer and Grant 2017 SGR 705); Spanish Government (Project RTI2018-095094-B-C21 “CONSENT”). The authors are with the UNESCO Chair in Data Privacy, but the views in this paper are their own and do not necessarily reflect those of UNESCO.

✉ Josep Domingo-Ferrer
josep.domingo@urv.cat

Alberto Blanco-Justicia
alberto.blanco@urv.cat

[1] Department of Computer Engineering and Mathematics, CYBERCAT-Center for Cybersecurity Research of Catalonia, UNESCO Chair in Data Privacy, Universitat Rovira i Virgili, Av Països Catalans 26, 43007 Tarragona, Catalonia

Introduction

Our society is increasingly dependent on digital technologies. Almost every social and economic activity of the developed and less developed countries has undergone some kind of digital transition. At least the following activities depend, at different levels, on information and communication technologies (ICT): governance, including elections, tax services, justice administration and communication with the citizens; business management and operations, including logistics, personnel and almost all day-to-day activities; banking transactions and the stock market; electronic commerce and all kinds of communication among people, such as email, instant messaging and social media; healthcare, in particular implants and telemedicine; and cars, that are moving from assisted driving to partial and ultimately fully automated driving. Therefore, it is of paramount importance that all actors in society can trust the technologies involved in the digital transformation.

In extremely general terms, one can define trust as "an expectation about a future behaviour"(Bamberger 2010), which would even include expecting the trustee to harm the truster. However, more usual acceptations of trust involve expecting a "good" behavior: according to New Oxford (2015) trust is "a belief in the reliability, truth, ability or strength of someone or something"; according to Robinson (1996), trust are "expectations, assumptions or beliefs about the likelihood that another's future actions will be beneficial, favorable or at least not detrimental". In the cybersecurity context, what the public can expect from a software system, service provider or governmental agency is that it behaves according to an established (sometimes implicit) agreement and that the actions by the trusted party do not cause the truster any harm, be it physical, economic or moral. Thus, cybersecurity is essential to build and sustain the trust relationship. The goal of cybersecurity measures is to ensure the availability and integrity of resources, as well as the confidentiality of stored data in the presence of threats such as natural disasters, human errors, corporate espionage, criminality, government-driven attacks, surveillance, terrorism, hacktivism, etc.

While the implementation of cybersecurity measures can be considered as mandatory, the specific techniques and the way they are applied can put at risk other important values, such as privacy, fairness and autonomy. If these values are not taken into account, the trust of the public in ICT and online services may well be marred, thereby potentially causing a negative impact on economy and society.

We analyze the synergies and tensions between the main ethical values to be preserved in a cybersecurity context, namely security, privacy, fairness and autonomy. From this analysis, we derive a value graph, and then we set out to identify the Hamiltonian paths in the graph that visit all the four aforementioned values, by leveraging their synergies and avoiding their tensions. Each path indicates a sequential approach to satisfying all values. For the sake of practicality, we illustrate our conceptual discussions with examples of enabling technologies. Our methodology can also be generalized for application to other domains and settings where several values that are partially synergetic and partially conflicting must be satisfied.

The rest of this paper is organized as follows. In "Relevant Values in Cybersecurity" section, we examine and define the main values in the context of cybersecurity, namely security, privacy, autonomy and fairness. In "Interplay Among Security, Privacy, Fairness and Autonomy" section, we construct a graph relating the above values by analyzing how they help or harm each other. In "Itineraries for Ethical Cybersecurity Based on the Value Graph" section, we identify the Hamiltonian paths in the graph and we explain and illustrate them in detail. Finally, in "Conclusion, Caveats and Generalization" section, we gather conclusions, caveats and methodology generalizations.

Relevant Values in Cybersecurity

We focus on four main values that are relevant in a cybersecurity scenario: security, privacy, fairness and autonomy. Admittedly, *transparency* (openness to public scrutiny) or the closely related honesty/sincerity (freedom from deceit or untruthfulness) might be regarded as well as relevant values for cybersecurity (EU SAM 2017). Yet, rather than viewing transparency/sincerity as a traditional ethical value, we will consider it as a procedural value that can enable ethical values.[1] For transparency/sincerity to actually induce ethical values, these values must be prevalent in the society conducting the public scrutiny. Otherwise, transparency does not guarantee much.

In specific application domains, additional values can come into play. For instance, see Loi et al. (2019) for an analysis of cybersecurity ethics in healthcare, where values such as non-maleficence (not harming the patient) and beneficence (doing good to the patient) arise.

For the sake of generality, we will keep the spotlight on the four aforementioned main values, which are defined in more detail in the next subsections.

Security

Security is the state of being free from danger or threat (New Oxford 2015). In particular, *cybersecurity* is security in computer systems and networks, and it is defined as the state of being protected against the criminal or unauthorized use of electronic data (New Oxford 2015).

In addition to referring to the above-described state of protection, cybersecurity is also used to denote the measures taken to reach such a state. More precisely, in European Commission (2013) and EU SAM (2016) cybersecurity is defined as "the safeguards and actions that can be used to protect the cyber domain, both in the civilian and military fields, from those threats that are associated with or that may harm its interdependent networks and information infrastructure. Cybersecurity strives to preserve the

[1] Similarly, in their discussion about the principles/values related to artificial intelligence, Floridi and Cowls (2019) mention explicability/transparency as a "new enabling principle" to be added to what they call "traditional bioethics principles": beneficence, non-maleficence, autonomy and justice/fairness.

availability and integrity of the networks and infrastructure and the confidentiality of the information contained therein."

Privacy

Privacy is a fundamental right included in Article 12 of the Universal Declaration of Human Rights. However, if privacy is defined as "the state or condition of being free from being observed or disturbed by other people" (New Oxford 2015) or "the right to be left alone" (Warren and Warren 1890), then complete seclusion is needed and this is hardly compatible with information technology and the information society. What is more, seclusion in the online world seems a *contradictio in terminis*.

It is more realistic to define privacy as "the right of an individual to decide what information about himself should be communicated to others and under what circumstances" (Westin 1970). Thus, rather than on non-disclosure, the emphasis should be on the ability of an individual or a group of individuals to *control and select* the disclosure of sensitive or confidential information about themselves.

Fairness

Fairness is the impartial and just treatment or behavior without favoritism or discrimination (New Oxford 2015). This implies an absence of bias and the assurance of equal treatment of individuals or groups of individuals by institutions, companies, etc. Fairness stems from the premise that equals should be treated equally, since according to the Universal Declaration of Human Rights, all human beings are born free and equal in dignity and rights (UN 1948). While there may be individual features, such as professional skill, health condition, criminal record, etc., that justify treating persons differently—a just treatment is one giving each person what he/she deserves—, other features, such as gender, race, religion, sexual orientation, etc., cannot be used to discriminate among people. Non-discrimination on certain attributes is legally mandated in many countries.

Autonomy

According to New Oxford (2015), autonomy is the state of being free from external control or influence. However, being completely free from any control or influence may be next to impossible when living in society.

A more realistic definition is to view autonomy as self-determination, that is, as the capacity of an individual to make an informed, uncoerced decision. In this sense, autonomy can be quantified as the boundaries of the actions individuals can take.

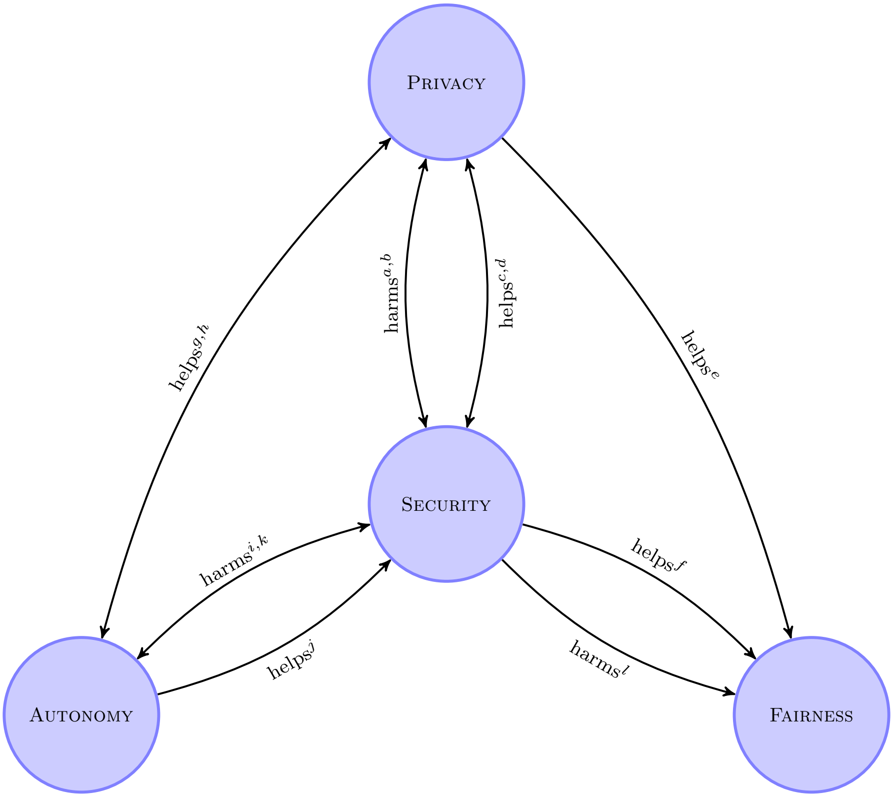

Fig. 1 Interrelationships among security, privacy, fairness and autonomy. Letters in edge labels are parenthetically included in subsection headings describing the corresponding relation

Interplay Among Security, Privacy, Fairness and Autonomy

In this section, we will examine the interrelationships among security, privacy, fairness and autonomy. They are graphically summarized in terms of help/harm relations in Fig. 1. A "helps" directed edge from one value to another means that the former value may help to achieve the latter value. In contrast, a "harms" directed edge from one value to another means that the former value may hamper achieving the latter value. In some cases there are synergies and in other cases tensions arise.

As the attribution of "helps" and "harms" edges between pairs of values requires expert knowledge, the following subsections describe the rationale behind each directed edge in the graph. Each subsection title contains the letter that labels the corresponding edge.

Security can Harm Privacy (a)

Network monitoring activities conducted by either governmental agencies or private security providers are normally viewed as a major clash between security and privacy. Specifically, monitoring by the state (e.g. police or agencies) is often regarded as sacrificing privacy in the altar of security. However, the extent to which such a sacrifice is real strongly depends on what is exactly monitored and under which conditions. Different degrees of monitoring can be distinguished:

1. Full Internet monitoring by an agency such as it is being discussed in Germany (AFP 2018) or is practiced by large agencies in China (Ma 2018) and the U.S. (the National Security Agency in the U.S. case);
2. Selective monitoring of offenders;
3. Monitoring of metadata that are not directly related to people, such as Internet protocol (IP) addresses, domain names that are related to criminal activities, etc.

The above argument shows that describing the relationship between security and privacy as a mere conflict is an oversimplification. Whereas the first case above (full Internet monitoring) can harm the privacy of many people, especially innocent people, the second and third cases (selective monitoring) can be controlled more easily by a system of checks and balances. It is crucial that every measure by the state be balanced between the desired effect of increasing the security of citizens and their right to privacy. Such checks and balances must not only be analyzed in the present but for the future as well; that is, data collected under a more rigorous law should not be used for other purposes if the law is changed.

Privacy can Harm Security (b)

Complete anonymity and secrecy of communications can be exploited by malicious entities to attack services without being discovered. One example are anonymization services like (Tor 2019), which offer the possibility to access websites and online services without disclosing one’s IP address. Although these services protect the privacy of users, they also pose a threat to the security and trustworthiness of online services, for instance because malicious activities cannot be traced back to the perpetrator and perpetrators may use the anonymizer to act with multiple identities (Sybil attacks).

Security can Help Privacy (c)

No matter whether privacy aids or hampers cybersecurity efforts, it is clear that cybersecurity to ensure data protection is needed to achieve any level of privacy. In particular, privacy is endangered whenever integrity and confidentiality—two basic security goals—are violated.

Furthermore, the third case of "Security can Harm Privacy (a)" section above (monitoring metadata) can be potentially beneficial for the privacy of citizens in some situations. For example, if an infected device is detected by a security agency and the user is warned, the state contributes to the privacy of this user, as otherwise the attacker could have had access to all the victim's data—confidentiality would have been violated.

Privacy can Help Security (d)

Let us consider how privacy and research on privacy help cybersecurity. Spear phishing attacks are phishing attacks directed at specific individuals. The availability of private information about individuals makes it easier for attackers to perform this kind of attacks, that may later lead to more critical attacks. On the other hand, strong and applicable privacy laws lead to better products also in terms of security. A good example is the discussion about electronic health dossiers that currently takes place in Switzerland (De Pietro and Francetic 2018) and many other European countries. The fact that privacy is needed and undisputed in such a sensitive environment leads to products that are engineered with better security.

Encrypting and/or anonymizing collected data are good ways to prevent an excessive invasion of the subjects' privacy, but the application of these techniques in the cybersecurity field is still neither well understood nor typically deployed. For example, the cybersecurity research roadmap (USDHS 2009) mentions the challenge of being able to share data on attacks with adequate privacy and, more generally, it identifies privacy-aware security as a hard problem in information security research. Nonetheless, the prize is worth the effort: if people and organizations feel their vulnerability data can be shared in a privacy-aware manner, more data will be shared, which will result in more efficient security countermeasures.

Privacy can Help Fairness (e)

When dealing with online services and the digital ecosystem, the discussion on fairness and non-discrimination is very closely related to privacy. For example, profiling services are arguably against the privacy of online users, and the typical consequences of such profiling activities can be most often related to an unfair treatment of the users. This connection between profiling and potential discrimination is made explicit in Article 71 of the European Union's General Data Protection Regulation (European Union 2016).

Therefore, privacy-enhancing techniques that reduce the amount or the granularity of profiling have beneficial effects on fairness.

Security can Help Fairness (f)

It is generally admitted that governments are not yet ready to provide a security level in the cyberspace that is comparable to the security they can offer in the physical space. Therefore, good security measures require private investment by companies and individual citizens. The lack of financial resources leaves part of the population and small companies vulnerable to attacks such as distributed denial of service (DDoS), extortion (e.g. ransomware), theft of confidential information, etc. It is important to note that the security community is very active in this matter and that large enterprises have taken as their responsibility to protect those who cannot protect themselves. A good example of this is Google's Shield Project (Shield 2019), which offers a free service to protect from DDoS attacks news sites and people who stand up for freedom of speech.

Another interesting topic is net neutrality, that is the principle whereby Internet regulators treat all transmissions (or packets) the same way, independently of their origin, destination and content. This principle is the basis for a fair Internet to everybody. Not only regulators, but also attackers can put net neutrality at risk. Interested parties can dominate big parts of the Internet, for example through the use of malware (botnets) to carry out illegal activities. Security measures are required to avoid these practices.

Autonomy can Help Privacy (g)

In "Privacy" section, we have defined privacy as the ability to control the disclosure of personal information. In fact, this is also known as *informational self-determination*, which is a right that was first mentioned in a German constitutional ruling (BVerfGE 1983) dated Dec. 15, 1983 in the following terms (English free translation):

> In the context of modern data processing, the protection of the individuals against unlimited collection, storage, use and disclosure of their personal data is encompassed by the general personal rights of the German Constitution. This basic right warrants in this respect the capacity of the individuals to determine in principle the disclosure and use of their personal data. Limitations to this informational self-determination are allowed only in case of overriding public interest.

In this light, thus, privacy is obtained as a result of the individual's autonomy. Without autonomy, there can be no privacy in the above sense.

Privacy can Help Autonomy (h)

If individuals know they are being observed and/or tracked, they are likely to self-censor, which clearly reduces their autonomy. In this respect, in many situations, privacy may be a precondition of autonomy. By connecting with the previous paragraph, we can close the circle by saying that *informational autonomy* is a necessary condition for *privacy* and, conversely, *privacy* is often a necessary condition for *autonomy of action*.

Autonomy can Harm Security (i)

In the physical space and in the cyberspace as well, the actions of an individual affect not only herself, but also the well-being and even the security of others. For example, when assessing the security of interconnected machines and networks, the security level of the whole network is typically the one of its weakest link. Therefore, to guarantee security for a community/system, responsibility and commitment to security are needed on the part of all individuals/components. This implies a restriction of autonomy.

As a matter of fact, if one wants to stay secure on the Internet, one has to stick to a set of best practices that impinge on one's autonomy. These best practices make the life of users more complicated due to higher complexity and poorer usability. For instance, they have to disable plugins, use a secure browser, use long passwords, multiple authentication factors, etc. People in general have taken responsible attitudes to fight cybercrime.

Autonomy can Help Security (j)

The Internet, referred to as the decentralized collection of autonomous interconnected networks, is managed by many individuals and organizations that act autonomously following their own interests. These autonomous parties have typically had good reactions to new security threats which would not have been possible without individuals taking the right steps. A good example of this is the rise of spam that led to very effective initiatives such as (Spamhaus 2019). Therefore, autonomous entities protecting themselves, and finding countermeasures to different threats, increase the security of the whole Internet. Open-source communities and non-profit organizations, on their side, work also to increase the security of all. A notable example is the Open Web Application Security Project (OWASP 2019), an online community that creates and publishes freely available articles, methodologies, tools, technologies and documentation for the protection of web applications. One of their most notable publications is the OWASP Top Ten, which is released regularly, and includes the ten most critical risks of web services and possible countermeasures. OWASP has become a *de facto* standard for web security.

User autonomy in disseminating information on vulnerabilities can also be security-beneficial. Being informed by users on vulnerabilities of their products is basic for vendors to craft countermeasures. However, vulnerabilities are typically not disclosed to the general public until a patch is available. While full disclosure to vendors only was the usual practice in the early days, the slowness and unwillingness of certain vendors to promptly fix their bugs and publish patches caused some users to exercise their autonomy and adopt immediate public disclosure, which may unfairly punish other vendors that would make an honest effort to handle their vulnerabilities. Hybrid disclosure, where the benign user does not announce the vulnerability knowledge to the public immediately, but instead allows the vendor some time to develop a patch, is a relatively common practice today. If the vendor does not release its patch before the deadline, the benign user informs the public about the vulnerability. Hybrid disclosure is a reasonable way for users to leverage their autonomy for the sake of security.

The zero-day vulnerability and exploit trade is currently legal, i.e. the autonomy of such dealers is deemed more important than security concerns. On the other hand, it is not clear whether security would improve or decrease if the vulnerability trade was regulated. We believe that this is an interesting conflict, whose resolution is not obvious.

Security can Harm Autonomy (k)

Some measures taken by organizations may impact the ability of autonomous entities to operate on the Internet. For example, it is perfectly legal to operate an e-mail server at home; no Request-for-Comments (RFC, the documents that contain the Internet specifications) or other standards forbid it. However, the reality is that it is next to impossible to do so because most organizations block the Simple Mail Transfer Protocol (SMTP, the Internet's e-mail protocol) traffic from home user IP ranges.

Governments also actively impose not-so-well justified measures that impact on the autonomy of companies. A good example are the National Security Letters issued by the U. S. Federal Bureau of Investigation (FBI), which force internet service providers to implement surveillance but hinder them from telling so.

Security can Harm Fairness (l)

Beyond restricting the autonomy of Internet users, restricting who can provide e-mail with the excuse of security is unfair: while the e-mail service was initially fully distributed, a small set of large companies are increasingly monopolizing its provision (Microsoft, Google, etc.) and individuals are *de facto* prevented from running their own e-mail servers.

Itineraries for Ethical Cybersecurity Based on the Value Graph

In previous sections, we have explained the relationships between security, privacy, autonomy and fairness depicted in the value graph of Fig. 1.

In this section, we assume that it is desirable to achieve all four values, which according to the previous analysis are not independent of each other. Therefore we will look for Hamiltonian paths[2] using only "helps" edges that visit all the graph nodes. (If only a subset of values was desired or if there were independent subsets

[2] In graph theory, a path in a graph G is a sequence of nodes of G such that an edge in G exists that connects each node to the next node in the sequence. A Hamiltonian path in G is a path that visits each node of G exactly once.

of values, subpaths visiting only the relevant subsets of nodes could be considered.) Clearly, fairness must be the last visited node in any path, because it only has incoming nodes. Thus we have the following possible Hamiltonian paths:

$$\textsc{Privacy} \overset{\text{helps}^{h}}{\Longrightarrow} \textsc{Autonomy} \overset{\text{helps}^{j}}{\Longrightarrow} \textsc{Security} \overset{\text{helps}^{f}}{\Longrightarrow} \textsc{Fairness} \tag{1}$$

$$\textsc{Autonomy} \overset{\text{helps}^{g}}{\Longrightarrow} \textsc{Privacy} \overset{\text{helps}^{d}}{\Longrightarrow} \textsc{Security} \overset{\text{helps}^{f}}{\Longrightarrow} \textsc{Fairness} \tag{2}$$

$$\textsc{Autonomy} \overset{\text{helps}^{j}}{\Longrightarrow} \textsc{Security} \overset{\text{helps}^{c}}{\Longrightarrow} \textsc{Privacy} \overset{\text{helps}^{e}}{\Longrightarrow} \textsc{Fairness} \tag{3}$$

The above three paths indicate three different itineraries to satisfy the four main values under study when aiming at cybersecurity. Interestingly, two among the three itineraries require to first meet intrinsically individual values (privacy, autonomy) in order to reach values that are more group-oriented (security, fairness).

A particular Hamiltonian path will be preferred if the technology solutions needed to implement its “helps” transitions are more easily deployable, more easily available, more cost-effective or better in any other relevant sense than those needed by the other Hamiltonian paths.

The following three subsections discuss in detail the steps of each of the three paths identified above.

The Privacy–Autonomy–Security–Fairness Path

If the path of Expression (1) is chosen, following the “helps” transitions and avoiding the “harms” transitions along this path in Fig. 1 yields requirements to be met to reach the four values.

The privacy–autonomy step. Privacy ought to be satisfied in such a way that:

- Transition “helps^{h}” facilitating autonomy applies. This means that users should not be observed or tracked, so that they do not feel drawn to self-censorship.
- Transition “harms^{b}” hampering security does not apply. In case of cybercrime, there should be mechanisms to trace what happened and who was responsible for what.

According to the European Union General Data Protection Regulation (GDPR, (European Union 2016)), data protection must be implemented by design and by default. Privacy in this sense can be attained by following the so-called privacy-by-design strategies (Hoepman 2014; Danezis et al. 2015): *Minimize*, *Hide*, *Separate*, *Aggregate*, *Inform*, *Control*, *Enforce* and *Demonstrate*. The amount of personal information processed should be minimal, it should be hidden from plain view, the processing should be done in distributed fashion whenever possible, personal information should be processed at the highest level of aggregation with the least possible detail in which it is still useful, data subjects should be informed whenever personal information is processed, they should have agency over the processing of their personal information, a privacy policy compatible with legal requirements should be enforced, and the data controler should be able to demonstrate compliance with the privacy policy and legal requirements.

Some privacy technologies go a step beyond and preclude security damage through privacy abuse. An example is anonymous electronic cash secure against double spending (Brands 1994): the users can anonymously spend their electronic cash (whereby privacy facilitates autonomy of spending), but if they spend the same cash twice, they lose their anonymity (whereby privacy cannot harm security). Another example are e-voting systems offering incoercibility, like Riera-Jorba and Castellà-Roca (2007): voters can anonymously cast their vote (whereby privacy favors the autonomy of voters), but they cannot prove to a third party what they voted (whereby the voters cannot sell their vote or be coerced to vote for a certain option).

The autonomy–security step. Autonomy should be satisfied so that:

- Transition "helps^j" facilitating security applies. Open and free collaboration of users should be encouraged to detect security threats and exchange information about vulnerabilities.
- Transition "harms^i" hampering security does not apply. Best practices should be disseminated and assistance should be given so that the autonomy of users and organizations does not result in their adopting poor security standards that might harm others.

Enabling technologies for autonomy to enhance or at least not to worsen security were mentioned in "Autonomy can Help Security (j)" section above. They are predicated on open communities or national agencies offering platforms for collaboration and exchange, and disseminating best practices. National agencies have the advantage of commanding more trust and hence they can even offer routine penetration testing to users that want to see how secure are their systems.

Peer-to-peer exchange of information on attacks and vulnerabilities may also strengthen security. However, the exchanged information must be privacy-protected for peers not to disclose information about their own weaknesses. In this respect, see "Privacy can Help Security (d)" section above and "The Autonomy–Privacy–Security–Fairness Path" section below.

The security–fairness step. Security should be satisfied so that:

- Transition "helps^f" facilitating fairness applies. National security agencies and/or large internet companies should take as their responsibility to protect those that cannot protect themselves. Furthermore, they ought to guarantee net neutrality. In that respect, recent moves the the U.S. Federal Communications Commission (Fung 2018) to end net neutrality go against fairness and even against security: some stakeholders can dominate large parts of the internet and attacks on those stakeholders can disable the internet to a large extent.
- Transition "harms^l" hampering fairness does not apply. Individuals and organizations should not be prevented from running their own internet services (e.g. e-mail) unless there is a clear security justification for such restrictions.

- Transition “harms^k” hampering autonomy does not apply. If in this path autonomy facilitates security, it would not make sense for security to lean on reducing autonomy. The rationale is to increase security by free and uncensored collaboration of users and organizations. In other words, security is attained bottom-up, rather than top-down.
- Transition “harms^a” hampering privacy does not apply. Achieving security by restricting privacy undermines autonomy, which in turn precludes the aforementioned bottom-up approach to security.

Regarding fairness, we have already dwelt on fair access to resources, such as the net or protection technologies. Other areas posing fairness problems are interactions (protocols) and automated decisions (such as loan granting, personel selection, insurance premium computation, etc.).

Enabling technologies to ensure fairness in protocols can be found in cryptography. As an example, in a cryptographic commitment scheme one party commits to a message at present, which is going to be revealed and verified in the future. On the one hand, others cannot determine the message content until the initiating party releases the cryptographic key that was used to create the commitment. On the other hand, the initiating party cannot alter the message content after having committed to it. Such schemes allow constructing fair protocols where no party can cheat. Based on fair protocols, one can build more complex secure protocols, such as e-voting protocols, certified mail exchange, secure contract signatures without notaries, etc.

As to fairness in automated decisions, anti-discrimination sanitization provides mechanisms to prevent classifiers from using “explanatory shortcuts” focusing on, say, ethnicity, religion, sexual orientation or any other attribute that the law forbids to discriminate on. Some sanitization mechanisms are based on pre-processing the training data using techniques akin to those of statistical disclosure control, but aimed at reducing the inherent bias in the data (Hajian et al. 2014). Others act directly on the automatically mined rules, either by eliminating some of them or by generalizing some of the conditions of these rules (Hajian and Domingo-Ferrer 2013; Hajian et al. 2015).

The Autonomy–Privacy–Security–Fairness Path

If one chooses the path of Expression (2), following the “helps” transitions and avoiding the “harms” transitions along this path in Fig. 1 yields the requirements to be met.

The autonomy–privacy step. Autonomy ought to be satisfied in a way that:

- Transition “helps^g” facilitating privacy applies. Autonomy allows informational self-determination and hence privacy.
- Transition “harms^i” hampering security does not apply. See “The Privacy–Autonomy–Security–Fairness Path” section above for comments on how to avoid this transition.

The privacy–security step. Privacy must be satisfied in a way that:

- Transition "helps[d]" facilitating security applies. Privacy-preserving exchange of information about attacks and vulnerabilities can make users and organizations more willing to share that kind of information, which results in enhanced security countermeasures.
- Transition "harms[b]" hampering security does not apply. See comments in "The Privacy–Autonomy–Security–Fairness Path" section above on how to avoid this transition.

Enabling technologies for privacy to help security consist of anonymization (in particular statistical disclosure control) and encryption techniques to protect the exchanged vulnerability and attack information. The most detailed release format is a microdata set, essentially a database table each of whose records carries the data corresponding to one entity (an organization or a user having been attacked, an attack, a vulnerability, etc.). While microdata sets can be extremely useful for security researchers, it is fundamental that their publication do not compromise anyone's privacy in the sense of revealing information about identifiable individuals.

Statistical disclosure control (SDC) is a data anonymization discipline that deals with the inherent trade-off between protecting the privacy of the individuals the data refer to and ensuring that the protected data are still useful to researchers. Several SDC methods have been proposed in the literature to protect microdata sets (Hundepool et al. 2012; Domingo-Ferrer and Mateo-Sanz 2002). Some of them aim to prevent identity disclosure (by making it difficult to determine the identity of the individuals to whom records correspond), whereas others try to avoid attribute disclosure (by making it difficult to estimate or bound the values of confidential attributes for specific individuals).

Beyond microdata sets and other statistical outputs, unstructured data can also be released for research, in the form of documents (for example, security logs, etc.). A way to anonymize documents is document redaction, that consists of removing or blacking out sensitive terms in the unstructured text of documents. Alternatively, one may resort to the more general technique of document sanitization (Bier et al. 2009), whereby sensitive terms are replaced (rather than removed) with generalizations (e.g., the last byte or bytes of IP addresses can be deleted). Document sanitization is more desirable than pure redaction, because it preserves better the utility of the protected output.

As to encryption, current techniques encompass symmetric and public-key cryptosystems, digital signatures, end-to-end encryption (which ensures that neither attackers nor service providers can read messages between users), homomorphic encryption (which allows for example using untrusted cloud service providers to compute on encrypted sensitive data), multiparty computation (to calculate functions on vulnerability/attack data held by different entities without revealing these data), and many other advanced functionalities.

The security–fairness step. This step coincides with the last step of the first path and it has been discussed in "The Privacy–Autonomy–Security–Fairness Path" section above.

The Autonomy–Security–Privacy–Fairness Path

If one chooses the path of Expression (3), following the "helps" transitions and avoiding the "harms" transitions along this path in Fig. 1 yields the requirements to be met.

The autonomy–security step. This step coincides with the second step of the first path and it has been discussed in "The Privacy–Autonomy–Security–Fairness Path" section above.

The security–privacy step. Security should be satisfied so that:

– Transition "helps^c" facilitating privacy applies. Security measures need to be taken to guarantee the integrity and the confidentiality of information on individuals and organizations, because this is essential for privacy preservation. In other words, if hackers can decide what is known about an individual or organization, the latter enjoys no informational self-determination and hence no privacy.
– Transitions "harms^a", "harms^k" and "harms^l" respectively hampering privacy, autonomy and fairness do not apply. How to avoid these transitions has been explained in "The Privacy–Autonomy–Security–Fairness Path" section above when discussing the security-fairness step.

The privacy–fairness step. Privacy should be satisfied in such a way that:

– Transition "helps^e" facilitating fairness applies. Privacy measures aimed at minimizing profiling, especially profiling based on sensitive attributes such as religion, ethnicity, sexual orientation, etc., have a positive impact on the fairness of automated decisions made on people.
– Transition "harms^b" hampering security does not apply. See comments in "The Privacy–Autonomy–Security–Fairness Path" section above on how to avoid this transition.

Privacy technologies that can facilitate fairness include putting the user in control of the profiling granularity or even making the user's profile available to the service provider only in encrypted form. For an example of the first option, see the system described in Blanco-Justicia and Domingo-Ferrer (2016) where consumers can control how much they are profiled by loyalty programs. An example of the second option is privacy-preserving implicit authentication (Domingo-Ferrer et al. 2015; Blanco-Justicia and Domingo-Ferrer 2018), where the service provider authenticates a user by comparing its encrypted features against an encrypted profile.

If the users cannot control the amount of profiling on them, then anti-discrimination technologies must be used by the service provider to ensure that automated decisions will not discriminate people by sensitive attributes that the law defines as non-relevant. See "The Privacy–Autonomy–Security–Fairness Path" section above for further remarks on anti-discrimination technologies.

Conclusion, Caveats and Generalization

Enough countermeasures exist to mitigate or even remove most of the current cybersecurity threats. While most countermeasures entail potential conflicts with ethical values other than security, giving up security would harm these values even more. Solutions should be assessed in terms of their effectiveness for prevention, detection or reaction against threats, and minimize their negative impact on ethical values different from security.

In this article, we have examined the interplay between security and three main additional values, namely privacy, fairness and autonomy. Due to their interrelation, it does not make sense to pursue any of those values in isolation. Our approach based on a value graph allows identifying all possible paths to satisfy all the considered values by resting on their synergies and avoiding their negative effects against each other.

In particular, we have identified three paths, given by Expressions (1), (2) and (3). A common feature of the first two paths is that the first two visited values (privacy, autonomy) refer to individuals, whereas the last two (security, fairness) have a more social dimension. In contrast, the third path interleaves individual and social values.

Caveats

To achieve all the considered values, the Hamiltonian paths identified on the value graph of Fig. 1 must be traversed along the "helps" edges, while avoiding the "harms" edges. The ability to do so depends on the availability of enabling technologies. In "Itineraries for Ethical Cybersecurity Based on the Value Graph" section, we have given examples of enabling technologies based on cryptography, anonymization and anti-discrimination sanitization.

Yet, cryptography, anonymization and anti-discrimination are also fraught with concerns, which we next summarize:

- Cybersecurity experts and software systems designers often lack sufficient knowledge of the power of advanced encryption, anonymization and anti-discrimination methods to improve cybersecurity systems and the protection of the consumers' rights and values.
- Governments are unconfortable with encryption technologies and they have tried to limit their imports and exports, the key sizes, etc., as reflected in the Wassenaar Arrangement (Wassenaar 1995) and local legislation (Koops 2013). In particular, governments use the anti-terror fight to argue against end-to-end encryption. Illustrative examples are the FBI-Apple dispute after the San Bernardino terrorist attack (Nakashima 2016) and the proposal by the United Kingdom's former Prime Minister David Cameron to ban end-to-end encryption after the Charlie Hebdo attack (Warren 2015). Proponents and defenders of such limitations on encryption, however, seem to fail to understand that criminal groups can develop their own encrypted messaging services and distribute them among their mem-

bers, leaving as a consequence a less protected general population and equally protected powerful and/or criminal groups.

- The cryptographic community should do a better job at reflecting on the prospective ethical ramifications of their technologies when these are deployed (Rogaway 2015).
- The necessary reduction of information ensuing from anonymization and anti-discrimination sanitization may decrease the effectiveness of detecting attacks and vulnerabilities.

See CANVAS (2019) and Domingo-Ferrer et al. (2017) for more details on the ethical dilemmas faced by cybersecurity practitioners.

Generalization

The methodology presented in this paper can be generalized to tackle the problem of satisfying any set of values in any application domain other than cybersecurity. Its general steps are:

1. Identify the relevant values and their relations;
2. Build a graph with values as nodes and “helps” and “harms” edges;
3. Identify in the value graph the Hamiltonian paths formed by “helps” edges;
4. Determine the technologies/approaches that allow following each Hamiltonian path and avoiding any “harms” edges;
5. Select the most suitable Hamiltonian path based on the particular constraints on affordable costs, available technologies, usability, etc.

Following the above steps cannot obviate the need of expert knowledge on the particular application domain under consideration. This expertise is needed to analyze the effects of each value on the remaining values when enforced on the application domain. Also, selecting the most appropriate Hamiltonian path requires knowledge of the feasible solutions, costs, technologies and other domain-specific constraints. The advantage of the above steps is that they tell the expert what to look for and how to systematize it in order to effectively combine values.

References

AFP. (2018). German spies can keep monitoring internet hubs, court rules. *The Local.de*. https://www.thelocal.de/20180531/german-spies-can-keep-monitoring-internet-hubs-court-rules.

Bamberger, W. (2010). Interpersonal trust—Attempt of a definition. Scientific Report, Technical University Munich.

Bier, E., Chow, R., Golle, P., Holloway King, T., & Staddon, J. (2009). The rules of redaction: Identify, protect, review (and repeat). *IEEE Security & Privacy*, *7*(6), 46–53.

Blanco-Justicia, A., & Domingo-Ferrer, J. (2016). Privacy-aware loyalty programs. *Computer Communications*, *82*, 83–94.

Blanco-Justicia, A., & Domingo-Ferrer, J. (2018). Efficient privacy-preserving implicit authentication. *Computer Communications*, *125*, 13–23.

Brands, S. (1994). Untraceable off-line cash in wallet with observers. In *CRYPTO'93* (pp. 302–318). Berlin: Springer.

Bundesverfassungsgericht. (1983). BVerfGE 65,1 - Volkszählungsurteil. 15 December. http://www.servat.unibe.ch/dfr/bv065001.html. Retrieved September 22, 2019.

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J.-H., Le Métayer, D., Tirtea, R., et al. (2015). *Privacy and data protection by design-from policy to engineering*. Heraklion: European Union Agency for Network and Information Security.

De Pietro, C., & Francetic, I. (2018). E-health in Switzerland: The laborious adoption of the federal law on electronic health records (EHR) and health information exchange (HIE) networks. *Health Policy*, *122*(2), 69–74.

Domingo-Ferrer, J., Blanco, A., Parra-Arnau, J., Herrmann, D., Kirichenko, A., Sullivan, S., Patel, A., Bangerter, E., & Inversini, R. (2017). CANVAS white paper 4-technological challenges in cybersecurity. The CANVAS project. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3091942. Retrieved September 22, 2019.

Domingo-Ferrer, J., & Mateo-Sanz, J. M. (2002). Practical data-oriented microaggregation for statistical disclosure control. *IEEE Transactions on Knowledge and Data Engineering*, *14*(1), 189–201.

Domingo-Ferrer, J., Wu, Q., & Blanco-Justicia, A. (2015). Flexible and robust privacy-preserving implicit authentication. In *IFIP SEC 2015* (pp. 18–34). Springer.

EU Scientific Advice Mechanism. (2016). *Scoping paper: Cybersecurity*. High Level Group of Scientific Advisors.

EU Scientific Advice Mechanism. (2017). *Cybersecurity in the European digital single market*. High Level Group of Scientific Advisors, Scientific Opinion No. 2.

European Commission. (2013). *Cybersecurity strategy of the European Union: An open, safe and secure cyberspace*. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.

European Union. (2016). *General data protection regulation*. Regulation (EU) 2016/679. https://gdpr-info.eu. Retrieved September 22, 2019.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*. https://hdsr.mitpress.mit.edu/pub/l0jsh9d1.

Fung, B. (2018). The FCC's net neutrality rules are officially repealed today. Here's what that really means. *The Washington Post*.

Hajian, S., & Domingo-Ferrer, J. (2013). A methodology for direct and indirect discrimination prevention in data mining. *IEEE Transactions on Knowledge and Data Engineering*, *25*(7), 1445–1459.

Hajian, S., Domingo-Ferrer, J., & Farràs, O. (2014). Generalization-based privacy preservation and discrimination prevention in data publishing and mining. *Data Mining and Knowledge Discovery*, *28*(5–6), 1158–1188.

Hajian, S., Domingo-Ferrer, J., Monreale, A., Pedreschi, D., & Giannotti, F. (2015). Discrimination and privacy-aware patterns. *Data Mining and Knowledge Discovery*, *29*(6), 1733–1782.

Hoepman, J. -H. (2014). Privacy design strategies (extended abstract). In *IFIP SEC 2014* (pp. 446–459). Springer.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Schulte Nordholt, E., Spicer, K., et al. (2012). *Statistical disclosure control*. Hoboken: Wiley.

Koops, B. -J. (2013) Crypto law survey. Version 27.0. February. http://www.cryptolaw.org. Retrieved September 19, 2019.

Loi, M., Christen, M., Kleine, N., & Weber, K. (2019). Cybersecurity in health—Disentangling value tensions. *Journal of Information, Communication and Ethics in Society*, *17*(2), 229–245.

Ma, A. (2018). China has started ranking citizens with a creepy 'social credit' system—Here's what you can do wrong, and the embarrassing, demeaning ways they can punish you. *Business Insider*. https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4.

Nakashima, E. (2016). Apple vows to resist FBI demand to crack iPhone linked to San Bernardino attacks. *Washington Post*.

New Oxford American Dictionary. (2015). 3rd edition. Oxford: Oxford University Press.

OWASP—Open Web Application Security Project. (2019). https://www.owasp.org/index.php/Main_Page. Retrieved September 19, 2019.

Project Shield—Protecting news from digital attacks. (2019). https://projectshield.withgoogle.com/public/. Retrieved September 19, 2019.

Riera-Jorba, A., & Castellà-Roca, J. (2007). *Secure remote electronic voting system and cryptographic protocols and computer programs employed*. U. S. Patent No. 7,260,552.

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, *41*(4), 574–599.

Rogaway, P. (2015). The moral character of cryptographic work. IACR Cryptology ePrint Archive, Report 2015/1162. https://eprint.iacr.org/2015/1162. Retrieved September 22, 2019.

The EU H2020-700540 "CANVAS" project (2016–2019). https://canvas-project.eu.

The Spamhaus Project. (2019). https://www.spamhaus.org. Retrieved September 19, 2019.

The Tor Project. (2019). https://www.torproject.org. Retrieved September 19, 2019.

UN General Assembly. (1948). *Universal declaration of human rights*. https://www.un.org/en/universal-declaration-human-rights/. Accessed 22 Sept 2019.

U.S. Department of Homeland Security. (2009). *A roadmap for cybersecurity research*. https://www.dhs.gov/sites/default/files/publications/CSD-DHS-Cybersecurity-Roadmap_0.pdf. Retrieved September 22, 2019.

Warren, L., & Warren, S. (1890). The right to privacy. *Harvard Law Review*, *4*(5), 193–220.

Warren, T. (2015) UK government could ban encrypted communications with new surveillance powers. *The Verge*. https://www.theverge.com/2015/1/12/7533065/whatsapp-imessage-ban-uk-government-encryption.

Wassenaar Arrangement. (1995) *The Wassenaar Arrangement on export controls for conventional arms and dual-use goods and technologies*. https://www.wassenaar.org. Retrieved September 19, 2019.

Westin, A. F. (1970). *Privacy and freedom*. New York: Atheneum.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.