Are Backdoor Mandates Ethical?— A Position Paper

Raphaël Khoury
Université du Québec à Chicoutimi

Sylvain Hallé
Université du Québec à Chicoutimi

Abstract—A growing number of jurisdictions have passed so-called backdoor laws, which mandate the inclusion of a vulnerability or weakness in software code, for the benefit of law enforcement agencies. In this position paper, we examine the trade-off involved in this scheme in the light of four different ethical systems, with particular emphasis on how academic research can inform the discussion.

INTRODUCTION

Over the past several years, several governments have, at times, pushed for the idea that commercial software should be required to include a ‘backdoor’, a deliberate vulnerability whose existence and exploitation mechanism is disclosed only to the appropriate authorities. This would enable the authorities to obtain access to the information contained in any device running this software when needed to react to criminal activity.

Notably, in 2018, Australia adopted the Telecommunications and Other Legislation Amendment (TOLA act) which mandates tech companies to provide law enforcement agencies with a way to decipher encrypted data stored by users on their systems. A similar bill was passed in the UK in 2016 and another one was proposed in the U.S. senate in 2020. Listing 1 provides a very simple—even naive example of how such a backdoor might be implemented.

The code merely checks if the username is equal to a hard-coded value, and automatically grants access if it is the case. Naturally, an actual backdoor will need to be much more cleverly hidden in the code to avoid detection.

In the past, proponents of this scheme held up the impossible promise of incorporating these backdoors in commercial tools without compromising end-user security in any way, a claim that most security professionals dismiss as an impossibility. The common retort being that it is impossible to leave a door open for the “good guys” while keeping it closed to the “bad guys”. In other words, any deliberate vulnerability introduced for the benefit of law enforcement could potentially be exploited by cybercriminals.

Indeed, there has already been at least one case of a vulnerability that many believe was deliberately inserted in software that was later exploited by malicious, possibly foreign ad-
versaries. The vulnerability in question is a weakness in a NIST-published random number generator[3].

Listing 1. A very simple backdoor

```c
gets(username);
gets(password);
if(username == "SecretValue")
    return true;
else
    login(username, password);
// proceed with regular login
```

However, in a 2019 speech at Fordham University, U.S. attorney general William Barr recasted the argument as a trade-off between the benefits and shortcomings of backdoors 1. The inclusion of backdoors, he freely admitted, does degrade the security of the end-point user. However, this minimal degradation is the price to pay for the large strides that backdoors can provide in combating terrorism and countering other classes of crime. The argument is best stated in the attorney general’s own words:

“All systems fall short of optimality and have some residual risk of vulnerability […].”

If one already has an effective level of security say, by way of illustration, one that protects against 99 percent of foreseeable threats, is it reasonable to incur massive further costs to move slightly closer to some theoretical optimality and attain, say, 99.5 percent level of protection[…]. Here, a company would not invest its own money to gain that kind of incremental benefit and society should not be asked to pay that cost to accomplish the same purpose.

Now, some argue the best way to achieve this slight incremental improvement is worth the cost of imposing those costs on society in the form of degraded public safety. I think this is untenable – again using a crude illustration, if the choice is between a world where we can achieve a 99 percent assurance against cyber threats to consumers, while still providing law enforcement 80 percent of the access that it requires […] or a world where we have boosted our cyber security to 99.5 percent for consumers but at a cost of reducing law enforcement’s access to zero percent – the choice for society should be clear.”

In other words, Attorney General Barr is arguing that we should accept the small degradation in the information security of end-users brought about by mandatory backdoors, in exchange for the large gains in national security against threats such as terrorism that this tradeoff will provide.

This formulation is much more useful since it takes the form of a trade-off, a concept familiar to security practitioners. Any security mechanism is in some way a trade-off, with costs and benefits, risks and drawbacks; and a large literature on risk analysis informs us on how to balance this trade-off in a reasoned manner. The formulation also has the benefit that it can be stated in a rather straightforward manner as an equation. Barr argues that backdoor are acceptable as long as:

\[ B_S > C_u \] (1)

where \( B_S \) is the cost to society and \( C_u \) is the aggregate cost to users.

In this position paper, we discuss the tradeoff that arises when the state mandates that software companies deliberately insert a vulnerability, termed a backdoor, into any software or device that performs encryption or allows secure communication with another principal. We eschew legal aspects, which are discussed elsewhere, approaching this issue from the perspective of ethics and focusing on how recent academic research can inform a reflection on the ethical aspects of the discussion. We also omit any technical discussion of the specific manner by which the backdoor could be implemented, only supposing the existence of a mandatory vulnerability present in the software.

---

1The text of his address is available at:
www.americanrhetoric.com/speeches/
williambarrycybersecuritykeynote.htm
In the remainder of this paper, we examine the question of mandatory backdoors in the context of different ethical postures, namely Utilitarianism, Kantian Ethics, Black Swan avoidance, and Social Contract theory. In each case, we consider how recent academic research can enrich the discussion. We find that each ethical system allows us to refine the above equation, and provide actionable advice about the creation of ethically acceptable backdoors.

Utilitarianism

From a purely utilitarian perspective, the trade-off would seem to be worthwhile. After all, how can we demand perfection in our protection against a particular class of attacks, namely cybercrime, at the cost of a large reduction in our protection against several other classes of crime? However, on closer inspection, the optimistic assessment of the trade-off seems to rest on a number of assumptions that may not bear out in practice.

In particular, one can only speak of a small degradation in the security level of the end-user, and of a comparatively large gain in societal security, brought about by the backdoor if knowledge of the underlying vulnerability remains confidential. If it were ever discovered and disclosed, a patch would presumably be issued, and the benefits of the backdoor will evaporate. The possibility that a malicious adversary could exploit the backdoor to further his nefarious goals also alters the costs-benefit calculus. The success of the scheme advocated by backdoor proponents thus hinges on the possibility that the vulnerability will remain undiscovered indefinitely.

Recent research however, casts doubt on the feasibility of keeping secret to the actual functioning of the backdoor. Indeed, early academic models on vulnerability discovery posited an infinite number of vulnerabilities, each with an equal probability of being discovered and exploited [18]. Since then, a large body of research has shown that some vulnerabilities are more likely to be discovered and exploited than others, and that software development practices will impact the number of vulnerabilities in code.

But could the code be so cleverly designed that the vulnerability is never discovered? A recent study by Clark et al. [6] shows that factors unconnected to the quality of the software itself play a large role in vulnerability discovery, with the amount of time that has elapsed since the release of the code—seen as a proxy for the familiarity of the adversary with the code, being particularly predictive. Indeed, the same phenomenon has been observed in cryptographic algorithms, despite their complexity and maturation level [5]. These findings seem to indicate that it may be impossible to ensure that the vulnerability will remain undiscovered indefinitely, rendering discussion of the comparative benefits and costs of the trade-off moot.

This question intersects with another topic of active recent research, that of vulnerability rediscovery, a phenomenon by which a vulnerability that has come to the attention of a group of researchers is quickly rediscovered independently by a different group of researchers. If the fact that a vulnerability has been discovered indicates that it is likely to be rediscovered in short order, then it will be that much harder to the authorities to maintain the secrecy of the backdoor.

Herr et al. recently examined multiple datasets and found that between 15% and 20% of vulnerabilities are rediscovered in the time frame between the moment a vulnerability is initially discovered and the moment it is made public [12]. This constitutes a rather narrow time span, and may plausibly understate the expected rediscovery rate in cases where a vulnerability is discovered and kept secret for an indefinite period of time.

Anecdotal evidence lends support to the hypothesis of frequent rediscovery. For instance, the Spectre and Meltdown vulnerabilities mentioned above were discovered separately and independently by four different groups of security researchers [8]. Likewise, a serious vulnerability in the gLib library was discovered independently in the span of a few months by at least three groups of researchers, including researchers at

\footnote{In particular, code analysis tools might be designed to detect specific types of vulnerabilities only.}

\footnote{The authors later revisited the question, and lowered their reported rate of vulnerability rediscovery in some cases [11].}
Google and at Red Hat Linux[7]. These simultaneous discoveries are made all the more striking by the large span of time that separates the introduction of these vulnerabilities from their discovery namely 20 years in the former case and 8 years in the latter one.

In this respect, it is also interesting to stress that discovered vulnerabilities often go unexploited. There are several reasons for this. Notably, recent research indicates that vulnerabilities for which an exploit code is difficult to create, or which the impact of the exploitation is limited are less likely to be exploited even after the public divulgence of the vulnerability[14]. The prospect of avoiding blackhat exploitation of the vulnerability even after it is discovered and disclosed mitigates the risks incurred by the creation of the backdoor.

The above discussion allows us to refine equation 1. From a utilitarian perspective, it can be argued that backdoors are acceptable if

$$B_S > p_d * p_e \sum_{u=1}^{n} C_u$$

(2)

where $p_d$ is the probability that the backdoor will be discovered, $p_e$ is the probability that it will be exploited if discovered, and $\sum_{u=1}^{n} C_u$ is the sum of the individual cost of the exploitation of the backdoor for each user.

This formulation points to specific steps that can be taken to render the backdoor more ethically acceptable. Notably, drawing upon recent research on this topic can aid in the creation of a backdoor that is less likely to be discovered and exploited, thus minimizing the right-hand side of equation 2.

**Kantian Ethics**

So perhaps the trade-off should be rejected? This is, after all, the most commonly shared opinion in the community of security professionals. It is also the conclusion one would reach by reasoning from Kantian ethical notions of categorical imperative. It is also possible to see in this stance an echo of the NSPE engineering code of ethics⁴, and its obligation to “avoid all conduct or practice that deceives the public.”, to “hold paramount the safety, health, and welfare of the public” and to be guided “by the highest standards of honesty and integrity”. If the trade-off provides no benefits to civil security, and only drawbacks to cybersecurity, then deliberately incorporating vulnerabilities in code seems to be at least deceitful, even reckless.

But are we absolutely certain that, in the absence of an effective trade-off of the type suggested by William Barr, the deliberate insertion of vulnerabilities in code provides no security benefits to the end-user?

Here, the question we seek to answer intersects with one of the most intriguing dimensions of computer security research, namely the human element of security, and the peculiar manner in which humans react to incremental security measures. A rich academic literature exists on this topic. Surprisingly, researchers have found that people often compensate the incremental addition of security measures with the adoption of riskier behavior. Conversely, the greater exposure to risk can lead an individual to act in a more prudent manner[1]. This phenomenon has been observed in areas as varied as mandatory bicycle helmets and car seat belts (which cause cyclists and motorists to adopt a more aggressive riding behavior) and the introduction of methane-proof lamps in mines in the 19th century (which induced miners to remain in the mines despite the suspected presence of methane leaks, with often fatal results).

More pertinent to the issue at hand is a study of data leaks in medical facilities by Miller and Tucker, who found that institutions that implemented data encryption saw an increase in reported data leaks, possibly because employees were more careless in their handling of sensitive data, drawn in a false sense of security that comes from the knowledge that the data was encrypted[15]. Informally speaking, individuals seem to have a “risk thermometer” and adjust their behavior by either increasing or decreasing their exposition to risk until they are comfortable with the risk level at which they operate.

This phenomenon intersects with another established observation in psychological research, namely that individuals tend to overestimate risks that are imposed on them, as opposed to risks to which they consciously choose to expose

⁴Available at : https://www.nspe.org/resources/ethics/code-ethics
If the finding of these studies hold, it would mean that the introduction of backdoors might counter-intuitively result in an overall improvement in end-user security provided it motivates at least some users to be more cautious in their use of information technology. For example, one of the most consequential decision that a user can make to improve his security posture is to avoid storing certain personal information datum on his device, nor to share them online. It is not completely unreasonable, — though not at all certain, that the awareness and discomfort of backdoors will lead users to adopt a more prudent behavior, and that the benefits of this prudence will outweigh any risk incurred by the vulnerability itself.

Other primordial steps needed to ensure the security of software, such as the diligent application of patches, are also in the hands of the user. If the inclusion of a backdoor pushes users in this direction, then this fact must be included in the trade-off, which we now restate as:

$$B_S + B_u > p_d p_e \sum_{u=1}^{n} C_u$$

where $B_u$ refers to the gains in the security of the end-user that are ultimately rooted in their reaction to the inclusion of the backdoor. While these gains may be difficult to quantify, academic research on the psychology of security can be used to craft the narrative in such a way as to maximize the equation above, thus rendering the trade-off more ethical.

Curiously, Miller and Tucker’s study was published a few years prior to a settlement agreement between a Boston hospital and the Massachusetts state government that mandated the hospital to use encryption technology in order to safeguard patient data. If the findings of this study hold, they reveal that the state has mandated a hospital to adopt a course of action that resulted in a reduction in data security for the patients, a striking illustration of how public policy choices can often have counter-intuitive repercussions.

This last conclusion is made even starker by another contemporaneous study: according to Choi et al., the occurrence of cybersecurity incidents in hospitals correlates with an increase in heart attack fatalities[4]. The authors blame this negative outcome on the increased complexity of the work environment following the introduction of cybersecurity countermeasures.

These surprising outcomes hint at the complexity, and even futility, of the type of predictive analysis that underpins the trade-off we seek to evaluate.

### Avoiding the Black Swans

Perhaps the truly ethical course of action is not to try to balance one risk against another, but rather to focus on avoiding the most serious categories of risks: the ones from which recovery can be difficult or even impossible. The basis for this code of ethics can be found in the writing of Nicholas Taleb [21], who argues that most of the hazards faced by many organizations are the result of black swans, rare events with devastating consequences, that are difficult to incorporate in risk analysis as commonly practiced (Figure 1). In particular, Taleb argues that such events cannot be predicted using the statistical tools normally employed in risk analysis, in part because the probability of occurrence of each particular black swan is so small.

The black swan theory has been the subject of renewed interest in the context of the current pandemic, that serves as a vivid reminder of the impact that rare and unpredictable events can wield.
The fear of a black swan is particularly salient in the context of information security, where a single adversary may attack thousands of targets simultaneously. Even more alarming is the prospect of an attack on physical or digital infrastructure, such as the Mirai botnet which shut down DNS servers by way of a distributed attack originating in the infection of hundreds of thousands of vulnerable IoT devices. Cognizant of this risk, information security experts have long advocated against the presence in systems of single points of failure: system components that, if successfully attacked, could bring about the failure of the entire system.

It could be argued that a backdoor, present in every software, and giving a potential adversary complete control over the vulnerable device, is the ultimate black swan of information security. A malicious adversary who discovers the vulnerability and devises a mechanism to exploit it may be in the position to inflict considerable damage on a multitude of individual users and businesses. There is also evidence that hackers seem to prefer high-value targets, despite the fact that successfully carrying out an attack against such a target requires sophisticated skills and considerable time.

When examining the issue under the lens of the black swan theory, the trade-off can be restated as:

\[ B_S + B_u > (p_d * p_e \sum_{u=1}^{n} C_u) + C_S \]  

(4)

where \( C_S \) refers to the costs to society that are incurred by a large scale ‘black swan’ type attack whose cost is borne by society in general, rather than by any specific user.

Another line of research points to a tentative solution to this problem, namely the study of software diversity. Drawing on an analogy to biological systems, researchers have argued that the robustness of systems could be improved if the program instance used by each user differs slightly from that of every other user. Researchers have even likened the current software environment as a monoculture in which a single malware can potentially infect every software instance. Slight variations between instances can thus limit the damages that an adversary can inflict with a single piece of malware. Researchers have also suggested that diversity may aid in the detection of malicious behavior.

If a requirement that each software include a backdoor is ever adopted, strategies drawn from research in diversity may be employed to lessen the risks incurred by a single point of failure. This could be done, for instance, by creating several different backdoors for each software, and randomly including one in each product instance.

Indeed, Nassim Taleb argues that the optimal mechanism for self-protection against the risks of black swans is to design systems that iteratively improve themselves when stressed, (a concept he calls antifragility). The process of searching for and patching vulnerabilities, thus incrementally improving the security of the underlying code, a practice that is clearly at odds with the inclusion of backdoors, is an elegant example of antifragility. This connection has already been made by researchers in the software engineering community.

Social contract theory

The final ethical system in the context of which we will examine the backdoor question is the social contract theory, espoused by Thomas Hobbes and others. This line of thought emphasizes the reciprocity between the civic obligations of the citizens and the services rendered by the state. In this view, the citizen of a modern state must accept to relinquish some of their rights, in exchange for the protection of the state and in order to the benefit from the services it confers. Hobbes and other philosophers have also argued for the need for proportionality between the obligations of the citizenry and those of the state. This vision does not introduce new variables to the equation presented above, but instead urges us to see the trade-off in a different, less competitive light.

We are intuitively familiar with this interplay, which is seen in the care given to wounded veterans and in the compensation paid for property.
seized through eminent domain. A more apropos example is the fact that the state indemnifies individuals who suffer rare side effects from mandatory vaccines. The reasoning being that since the state mandates that citizens must be vaccinated, then it is incumbent on the state to assume the costs incurred by the side effects of vaccines.

This situation is in many ways analogous to the one discussed in this paper, whereby an illicit exploitation of the backdoor by malicious adversaries is akin to a kind of side effect of its mandatory inclusion in the code—a side effect that the state may be seen as obligated to redress. This aspect of the question must not be neglected when evaluating the costs and benefits of the trade-off.

Unfortunately, in this respect, academic research is less able to provide actionable insights. Indeed, attempts to quantify the costs incurred by the victims of vulnerabilities and their attendant cyberattacks only highlight the level of uncertainty in this regard. To illustrate this situation, one has only to consider the large discrepancies in the estimates of the annual costs of cyberattacks in different studies. For example, the 2018 Norton report estimates the global cost of cybercrime at 172M USD [16] while a 2020 study by the firm McAfee place the global cost at up to 1 Trillion USD [20]. This uncertainty makes a formal risk analysis-based evaluation of the trade-off even more difficult. Moreover, the human toll of cyberattacks, which is even harder to quantify, must also be taken into consideration. In particular, the psychological and emotional impact of having one’s personal correspondence and one’s photographs made public following a data leak does not easily lend itself to a monetary characterization, but it is certainly not inconsequential.

But while theories of social contract can form the basis for a requirement for software companies to comply with a government mandate to insert a backdoor in their code, these theories also serve to circumscribe the behavior of the state in its interaction with the citizenry. In particular, the citizen’s ascent of backdoors may be contingent on a guarantee that it will only be used in a manner consistent with the stated objectives of the backdoor, namely the fight against terrorism and other classes of serious crimes. Even those individuals who are most sympathetic to law enforcement may be reluctant to accept backdoors if they are widely used to spy on common citizens.

A thorough examination of this question has an important legal and political component, and is beyond the scope of this paper. However, it is interesting to bring attention to the risk of “slippery slope”, whereby the context in which it is permissible to exploit a backdoor widens over time. In this regard, it is interesting to recall the dispute that arose between the FBI and Apple corp. over court orders that would have compelled Apple to unlock an iPhone 5c used by one of the perpetrators of the San Bernardino terrorist attack of 2015. The case was in many ways ideal from the perspective of the state: the phone belonged to the suspect’s employer, who assented to the search, the suspect had died, so privacy objections did not enter into consideration, and the target of the court orders elicited little sympathy from the public. The case was mooted before reaching a resolution since the FBI found a different mechanism to break into the phone, but it is instructive to note that while the case was making its way in the court system, the Justice Department was planning on using the precedent established that would be established by this case to unlock phones in nine other cases, involving mostly low-level drug crimes.

For many citizens, the fear of government encroachment on their privacy, rather than the fear of exploitation by malicious cybercriminals, may well be the main driver of the resistance to the tradeoff suggested by Barr.

Conclusion

In this position paper, we analyze the question of state-mandated backdoors, from the perspective of four ethical systems, namely Utilitarianism, Kantian ethics, Black swan avoidance, and Social contract. We do not take position on this delicate issue, focusing instead on how recent advances in academic research can shed light on the discussion. In particular, the different ethical postures we consider allows us to state the trade-off involved in the inclusion of backdoors in the form of an equation, whose variables are design choices of the backdoor on which current research provides valuable insights. More specifically, we argue that current research leads us to make the
following recommendations:
• the backdoor should be of a type of vulnerability that is less likely to be rediscovered by potential adversaries, and less likely to be exploited if discovered, thus maximizing the benefits of the backdoor;
• authorities must continuously monitor the impact that the introduction of the backdoor will have, in order to detect any unexpected outcome;
• introduce diversity, to minimize the risks incurred by a common point of failure;
• further research is also needed to quantify the cost of cyberattacks, the estimating probability of vulnerability rediscovery and predicting the ways users and adversaries may react to the introduction of the backdoor, before a risk-analysis can successfully conducted.

It should be mentioned however, that several important aspects of this question have been omitted from this paper. Notably, the question of whether privileged users, such as bankers and government officials, will have access to special “backdoor-free” instances (Willam Barr implied that this would be case in his address) has not been discussed. The related issue of which criteria determines who qualifies for a more secure instance of the software, has also not been raised. Furthermore, the existence of the backdoor (and the fact that its existence is public knowledge) may lead malicious individuals to simply eschew the use of certain technologies, nullifying any benefit to law enforcement.

REFERENCES
11. T. Herr and B. Schneier. What you see is what you get: Revisions to our paper on estimating vulnerability rediscovery, July 20017.
13. R. Khoury, A. Hamou-Lhadj, and M. Couture. A formal framework for evaluating the effectiveness of diversity when applied to security. In IEEE Symposium: Computational Intelligence for Security and Defence Applica-


