The U.S. Vulnerabilities Equities Process: An Economic Perspective

Tristan Caulfield[1], Christos Ioannidis[2], and David Pym[13]

[1] University College London
t.caulfield@ucl.ac.uk, d.pym@ucl.ac.uk
[2] Aston Business School
c.ioannidis@aston.ac.uk
[3] The Alan Turing Institute

Abstract The U.S. Vulnerabilities Equities Process (VEP) is used by the government to decide whether to retain or disclose zero day vulnerabilities that the government possesses. There are costs and benefits to both actions: disclosing the vulnerability allows the the vulnerability to be patched and systems to be made more secure, while retaining the vulnerability allows the government to conduct intelligence, offensive national security, and law enforcement activities. While redacted documents give some information about the organization of the VEP, very little is publicly known about the decision-making process itself, with most of the detail about the criteria used coming from a blog post by Michael Daniel, the former White House Cybersecurity Coordinator. Although the decision to disclose or retain a vulnerability is often considered a binary choice—to either disclose or retain—it should actually be seen as a decision about timing: to determine *when* to disclose. In this paper, we present a model that shows how the criteria could be combined to determine the optimal time for the government to disclose a vulnerability, with the aim of providing insight into how a more formal, repeatable decision-making process might be achieved. We look at how the recent case of the WannaCry malware, which made use of a leaked NSA zero day exploit, EternalBlue, can be interpreted using the model.

1 Introduction

Governments, for national security, military, law enforcement, or intelligence purposes, often require an ability to access electronic devices or information stored on devices that are protected against intrusion. One way this access can be achieved is through the exploitation of vulnerabilities in the device's software or hardware. To this end, governments acquire, through a number of different methods, knowledge of these vulnerabilities—which are usually unknown to the software vendor and users—and how they may be successfully exploited.

However, the role a government plays is dual: in addition to the national security and law enforcement purposes above, which may require the exploitation of vulnerabilities, the government is also responsible for defending its national

assets in cyberspace. It has a responsibility to protect its own government and military networks, the nation's critical infrastructure, as well as the information assets of its businesses and citizens. When a government acquires knowledge of a vulnerability, this dual role presents a conflict. The government must decide between two competing national security interests: whether to retain the vulnerability, keeping it secret so it can be used to gain access to systems for intelligence purposes, or if it should instead be disclosed to the vendor, allowing it to be fixed so that the security of systems and software can be improved.

In the United States, this decision is now guided by the Vulnerabilities Equities Process (VEP), which the government uses to assess whether to retain or release each vulnerability it acquires. Publicly, relatively little is know the criteria used in this assessment. A Freedom of Information Act request from the Electronic Frontier Foundation (EFF) saw the release of a redacted version of a document [4] that describes how the VEP works within the government, but without any indication of how the decision to retain or disclose is made. A blog post [5] in April, 2014 by Michael Daniel, the White House Cybersecurity Coordinator, provided some insight, revealing a number of factors that are used in the decision-making process, but also that 'there are no hard and fast rules'.

The factors listed in the blog post are very high-level concepts, describing *what* decision-makers consider, but not *how* they do so. For example, some of these factors describe values, such as 'the extent of the vulnerable system's use in the Internet infrastructure' or 'the risks posed and the harm that could be done if the vulnerability is left unpatched', and yet no there is no indication of how they can be quantified or compared against each other. Given the lack of hard and fast rules, it is not unreasonable to assume that decisions are made on an ad hoc, case-by-case basis.

There has been some discussion and commentary about the VEP—and about how vague known information about it is. A June 2016 discussion paper by Schwartz and Knake [24] examines what is publicly known about the VEP and makes a number of recommendations to improve the process. Among these is the recommendation to 'make public the high-level criteria that will be used to determine whether to disclose ... or to retain [a] vulnerability' and that it is possible to 'formalize guidelines for disclosure decisions while preserving flexibility in the decision-making process'. Similarly, a September, 2016 EFF blog post [3] recommends that the government be more transparent about the VEP decision-making process, including the criteria used, and that the policy should be 'more than just a vague blog post'.

This paper aims to further understanding of how the different factors that are used in a VEP decision can be included in a more formal decision-making process. The intent is not to be normative: we do not aim to say, for example, how much potential harm is an acceptable trade-off for the benefits gained from exploiting a particular vulnerability. Instead, we look at different possible ways in which each of the factors may be evaluated or quantified, how the factors relate to each other, and how this information could be combined to make a decision. Specifically, we present the government's decision about whether or

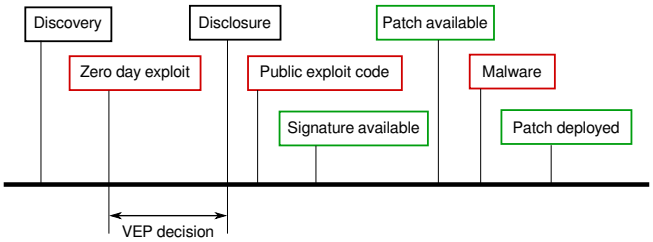

Figure 1. The vulnerability timeline, showing the events that can occur from the discovery of a vulnerability to its eventual patching. This is a guideline: not all of these events will occur for every vulnerability, and the order in which they occur may differ.

not to disclose a vulnerability as a timing problem, where the solution is the optimal amount of time to delay disclosing the vulnerability given the costs and benefits of doing so. We then look at how how the long delay before the disclosure of the EternalBlue vulnerability can be interpreted using this model.

Section 2, next, presents background information about vulnerabilities, exploits, and their increasing use by governments. Next, Section 3 introduces the VEP: its purpose and origins, the factors used in decision-making, and the discussion and debate surrounding the process. Section 4 examines each of the factors in detail, looking at how they might affect the disclosure decisions, and Section 5 looks at how the factors could be combined into a model to determine the optimal time of disclosure. Section 6 then looks at the WannaCry malware and the timing of the disclosure of the leaked vulnerability it used.

2 Background

During the development of a piece software, flaws—or *bugs*—may arise in its design or implementation which cause the software to behave differently than intended. A bug that can cause behaviour affecting the security of a system is called a *vulnerability*. An *exploit* is a technique or action (for example, a piece of software or a series of commands) that can be used to take advantage of the vulnerability. An *attack* is the use of an exploit to attempt this.

Creating software without bugs is a very difficult challenge and is not economically feasible for most software. As such, software often has to be updated after its initial release in order to fix bugs discovered later on. A vulnerability, once discovered, may be *disclosed*—either to the vendor directly, through an organization such as CERT, which coordinates disclosures with the vendor, or publicly. Once the developer is aware of the vulnerability, they may work to create a fix that removes the vulnerability. An updated version of the software containing the fix, called a *patch*, is then released; end-users of the software must then apply this patch to their systems to remove the vulnerability.

The sequence of events including the discover of a vulnerability, the creation of an exploit for it, the vulnerability's disclosure, and eventual patching is known as the vulnerability timeline. Figure 1, adapted from [2], illustrates this sequence

of events. Related to this is the notion of the window of exposure, discussed in [20], which is the time from the creation of an exploit until systems are patched during which systems are at risk from a vulnerability. This window can be reduced by improving the speed with which patches are produced and deployed. The VEP deals with the government's decision to disclose or retain *zero day* vulnerabilities; these are vulnerabilities that are unknown both publicly and to the software developer, so named because the developer and end users have had zero days to fix or mitigate the vulnerability. Disclosing the vulnerability allows a patch to be produced sooner, reducing the window of exposure.

The timeline includes an event for a signature becoming available, which indicates the the availability of methods to mitigate the vulnerability before the official patch has been released, including anti-virus or intrusion-detection signatures. There is also a distinction between the public release of exploit code and the development of malware. The former refers to code that utilizes the exploit—perhaps as a proof-of-concept or demonstration that the exploit works—but does not cause significant damage; the latter refers to more sophisticated and damaging uses of the exploit. However, publicly publishing proof-of-concept code can make it easier for more damaging exploits to be developed. All of the events in the timeline do not always occur for each vulnerability, and the ordering of events and the time between them is fluid. For example, there might not be a zero day exploit, or the patch could be released before any exploit is developed.

2.1 Increasing use of vulnerabilities

The exploitation of vulnerabilities by both governments and other parties is growing—and this is not surprising. The use of digital technologies in all aspects of life and business continues to grow at an astounding rate and more and more information is stored on electronic devices. Access to these devices and the information stored on them has value. For governments this could be the value of intelligence, the ability of law enforcement to conduct surveillance, or the ability to disrupt systems. For criminal actors, access to these systems can enable a host of different crimes, from theft or ransom of information to sabotage.

Evidence of the increasing importance of vulnerabilities to all parties is the rise of the market for vulnerabilities. A 2007 paper by Miller [12] documents the author's attempts to sell zero day exploits, which was both difficult to do and not extremely lucrative. A 2012 article in Forbes [8] gave a list of prices for zero day exploits, including a range of $30,000–$60,000 for Android, and $100,000–$250,000 for iOS—values that were surprising to Schneier at the time [22]. Today, the market is even more established, and exploits fetch a much higher price. Companies such as Zerodium buy exploits from security researchers and resell them to customers, including governments. Zerodium is currently offering researchers up to $200,000 for Android exploits and up to $1,500,000 for iOS exploits [25]. Other products that are less secure (so it is easier to find exploitable vulnerabilities) or less popular fetch lower prices. This increase in market price (and the expansion of the market itself) over the last decade is an indicator of the increasing demand for and importance of zero day exploits.

Often, exploits can be purchased with an exclusivity agreement, meaning that it will not be sold to anyone else. However, exclusivity agreements are no guarantee that the vulnerability will remain undiscovered by others. Other researchers, governments, or criminals may independently discover the same vulnerability. This is what causes the tension between the dual roles of the government: just because it it believes it is the only entity with access to a vulnerability does not mean it will not be used against assets it is charged to protect. Thus, every decision to retain a vulnerability instead of disclosing it so it can be fixed and patched increases the risk to systems the government aims to protect.

3 The Vulnerabilities Equities Process (VEP)

The Vulnerabilities Equities Process was created to address the tension between the offensive and defensive missions of the government. Schwartz and Knake [24] provide a thorough explanation of the background and origins of the VEP, and Healey [9] also gives a good overview; we will provide a brief summary here.

President George W. Bush signed a directive [14] in 2008 creating a government-wide Comprehensive National Cybersecurity Initiative (CNCI). This initiative required a number of government departments to develop a plan for coordinating the 'application of offensive capabilities to defend US information systems', which led to the production of the VEP document [4] in February, 2010.

A redacted version of the VEP document was obtained via a Freedom of Information Act request by the EFF. The document begins by stating its purpose:

> This document establishes policy and responsibilities for disseminating information about vulnerabilities discovered by the United States Government (USG) or its contractors, or disclosed to the USG by the private sector or foreign allies in Government Off-The-Shelf (GOTS), Commercial Off-The-Shelf (COTS), or other commercial information technology or industrial control products or systems (to include both hardware or software). This policy defines a process to ensure that dissemination decisions regarding the existence of a vulnerability are made quickly, in full consultation with all concerned USG organizations, and in the best interest of USG missions of cybersecurity, information assurance, intelligence, counterintelligence, law enforcement, military operations, and critical infrastructure protection.

The document also specifies conditions for whether or not a vulnerability is entered into the VEP: 'to enter the process a vulnerability must be both newly discovered and not publicly known' but that 'vulnerabilities discovered before the effective date of this process need not be put through the process'. The VEP document creates an Equities Review Board (ERB) that makes the decision about disclosing or retaining a vulnerability, establishes an Executive Secretariat, specifies how government agencies that come into possession of a vulnerability should notify the Executive Secretary, and how agency-designated

Subject Matter Experts (SMEs) hold discussions to evaluate the course of action for each vulnerability.

In short, the VEP document specifies how the process of submitting vulnerabilities works, how the various stakeholders have inputs, and how the process is managed. It does not mention what inputs or factors are used when making a decision, nor how any such factors would be considered.

3.1 The Daniel blog post

Information about the VEP was first released under the Obama Administration in 2014, in response to allegations by Bloomberg News [18] that the NSA was aware of and had exploited the Heartbleed vulnerability in OpenSSL, which the NSA denied. The White House commented, saying that the NSA would have disclosed the vulnerability, had they known about it, and in most cases would disclose any vulnerability discovered to allow it to be fixed. Referring to the VEP: 'unless there is a clear national security or law enforcement need, this process is biased toward responsibly disclosing such vulnerabilities' [19, 15].

Further information about the VEP came in the form of a blog post [5] by Michael Daniel, the White House Cybersecurity Coordinator, responding to the debate caused by the Heartbleed vulnerability. In it, Daniel discusses the trade-offs between disclosing and retaining a vulnerability— 'disclosing a vulnerability can mean that we forego an opportunity to collect crucial intelligence' but 'building up a huge stockpile of undisclosed vulnerabilities while leaving the Internet vulnerable and the American people unprotected would not be in our national security interest'.

Following this, Daniel provides the only public insight into the factors that are considered when deciding to retain or disclose a vulnerability:

> We have also established a disciplined, rigorous and high-level decision-making process for vulnerability disclosure. This interagency process helps ensure that all of the pros and cons are properly considered and weighed. While there are no hard and fast rules, here are a few things I want to know when an agency proposes temporarily withholding knowledge of a vulnerability:
> - How much is the vulnerable system used in the core internet infrastructure, in other critical infrastructure systems, in the U.S. economy, and/or in national security systems?
> - Does the vulnerability, if left unpatched, impose significant risk?
> - How much harm could an adversary nation or criminal group do with knowledge of this vulnerability?
> - How likely is it that we would know if someone else was exploiting it?
> - How badly do we need the intelligence we think we can get from exploiting the vulnerability?
> - Are there other ways we can get it?
> - Could we utilize the vulnerability for a short period of time before we disclose it?

- How likely is it that someone else will discover the vulnerability?
- Can the vulnerability be patched or otherwise mitigated?

These factors are weighed 'through a deliberate process that is biased toward responsibly disclosing the vulnerability'—but what this decision-making process is remains unknown.

3.2 Debate and recommendations

The VEP document and the Daniel blog post have been analysed and criticized a number of times. Schwartz and Knake [24] explore the history of the VEP and what is known about it from various sources and make recommendations to improve the process.

Several of the recommendations concern the decision-making process and are of interest here. First, 'the principles guiding these decisions, as well as a high-level map of the process that will be used to make such decisions, can and should be public'. Next, 'make public the high-level criteria that will be used to determine whether to disclose to a vendor a zero day vulnerability in their product, or to retain the vulnerability for government use'. Finally, if a vulnerability is not disclosed, the process should 'ensure that any decision to retain a zero day vulnerability for government use is subject to periodic review' and that vulnerabilities should be 'disclosed to the responsible party once (1) the government has achieved its desired national security objectives or (2) the balance of equities dictate that the vulnerability should be disclosed'.

The EFF also makes recommendations about the VEP. In August, 2016, an entity naming itself 'The Shadow Brokers' released a collection of files containing code for exploiting vulnerabilities in various firewall products from vendors such as Cisco and Fortinet. These exploits were linked to the NSA and, crucially, were exploiting previously unknown zero day vulnerabilities. The exploit code was stolen in 2013 and the NSA was aware it had been exposed, but the vulnerabilities were never disclosed.

In response to this, the EFF wrote in [3]:

> We think the government should be far more transparent about its vulnerabilities policy. A start would be releasing a current version of the VEP without redacting the decision-making process, the criteria considered, and the list of agencies that participate, as well as an accounting of how many vulnerabilities the government retains and for how long. After that, we urgently need to have a debate about the proper weighting of disclosure versus retention of vulnerabilities.

Similarly, Mozilla discusses the VEP in response to the Shadow Brokers leak [6] and makes recommendations, including:

- All security vulnerabilities should go through the VEP and there should be public timelines for reviewing decisions to delay disclosure;

- All relevant federal agencies involved in the VEP must work together to evaluate a standard set of criteria to ensure all relevant risks and interests are considered;
- Independent oversight and transparency into the processes and procedures of the VEP must be created. All security vulnerabilities should go through the VEP and there should be public timelines for reviewing decisions to delay disclosure.

Common to these three sets of recommendations is the desire for greater insight into the decision-making process and the factors or criteria that are used. Additionally, the recommendations from Schwartz and Knake and Mozilla are both concerned with the timing for reviews of vulnerabilities that have been retained. Proposed legislation, the Protecting our Ability To Counter Hacking (PATCH) Act [7], would turn the VEP into law and allows for periodic review of vulnerabilities—meaning that a vulnerability could be used for a time and then disclosed. We agree with these recommendations and, in the next two sections, we examine the factors from the Daniel blog post—to better understand how they might influence the decisions made—and then present a model for a decision-making process that utilizes the different factors to determine the optimal time for disclosure.

4 Factors

The first step in improving understanding of the decision-making process is to focus on the factors involved and try to understand in greater detail what they mean and how they can be measured. The next step is then to examine how they affect the decision. The choice to retain a vulnerability gives a benefit to the government: it allows the collection of additional information for national security, intelligence, or law enforcement purposes; it also brings a cost: the increased risk of harm to its own networks, businesses, and individuals. The government aims to find the correct balance between these two, and each of the factors affects the outcome of this decision.

As discussed above, the VEP has two possible outcomes. First, a vulnerability may be disclosed; if this is the case, then the process ends with the disclosure. The other outcome is the decision to retain the vulnerability for use. If this is the case, then according to the process, the decision should be reviewed again at some point in the future and either disclosed or retained further. The VEP can be seen, then, as a timing problem: given the costs and benefits associated with disclosing or retaining a vulnerability, when is the best time time to disclose?

Each of the factors in the decision-making process can then be considered to have either an accelerating or a retarding effect on the time of disclosure. For example, if a factor reduces the risks or costs of non-disclosure, it will tend to delay disclosure; if it increases the risks, then it will move disclosure forward.

In this section, we will discuss each of the factors from the Daniel blog post, looking at what they mean and how they can be measured, and examining their impact on the costs and benefits to the government.

4.1 Extent of use

How much is the vulnerable system used in the core internet infrastructure, in other critical infrastructure systems, in the U.S. economy, and/or in national security systems?

The meaning of this factor is straightforward, as is its measurement. Data about the number of units sold or deployed for a particular device or piece of software is not difficult to acquire or estimate. This factor is related to the risks and harm, below—where and how widely a device with a vulnerability is used will affect the potential risks and harms.

The extent of use may change over time. For example, end users might switch to newer devices or upgrade to newer versions of software that are not affected by the vulnerability.

Effect. This factor affects the decision to disclose in both directions, though not necessarily equally. First, a vulnerability in a widely-used device or piece of software can potentially cause harm to a larger group of individuals, businesses, or systems; this will have an accelerating effect on the time of disclosure. However, the opposite is also true: a vulnerability in a more widely-deployed system can potentially allow the government to access a greater number of systems, which would delay disclosure.

4.2 Risks and harm

Does the vulnerability, if left unpatched, impose significant risk? How much harm could an adversary nation or criminal group do with knowledge of this vulnerability?

There are many potential ways in which exploitation of the vulnerability by others could cause harm. At a national level, there are potential harms from the compromise of government networks or the disruption of critical infrastructure. For businesses, harms can include direct monetary loss (from fraud, theft, sabotage, or ransomware) or loss of competitive ability (from industrial espionage), and also reputational harm caused by a breach. Harms to individuals include, for example, direct losses from crime, identity theft, and loss of privacy.

For each vulnerability, the risks of each of these harms will be different—it is unlikely, for example, that a vulnerability in an industrial control system will present much risk of identity theft to individuals, but the same vulnerability could present a large risk to infrastructure or businesses. The government must estimate how likely different harms are for each vulnerability, as well as the magnitude of those harms; this is related to the extent of use: where and how much devices with the vulnerability are used will affect the likelihood and impact.

Effect. That the government considers risks and harms instead of simply losses implies that they distinguish between the risk of discovery and use of the vulnerability by others and the 'lumpiness' of the harm. If the government has an

aversion to substantial harm from single events, then its potential presence makes the decision to retain the vulnerability costlier, and will accelerate disclosure, even if the likelihood is low. If the risk of discovery and use is very high, even if the potential harm is modest in terms of impact on individuals or businesses, that will also accelerate the decision to disclose.

4.3 Detect exploitation by others

How likely is it that we would know if someone else was exploiting it?

This is hard to estimate without knowledge of the government's capabilities. A quote from [11] in the aftermath of the Shadow Brokers leak gives an indication that the NSA does have such an ability:

> After the discovery, the NSA tuned its sensors to detect use of any of the tools by other parties, especially foreign adversaries with strong cyber espionage operations, such as China and Russia.
> That could have helped identify rival powers' hacking targets, potentially leading them to be defended better. It might also have allowed U.S officials to see deeper into rival hacking operations while enabling the NSA itself to continue using the tools for its own operations.
> Because the sensors did not detect foreign spies or criminals using the tools on U.S. or allied targets, the NSA did not feel obligated to immediately warn the U.S. manufacturers, an official and one other person familiar with the matter said.

Effect. If the government has a high confidence in their ability to detect the exploitation of the vulnerability by others then this will have a delaying effect on disclosure time. From the quote above, this appears to be the case. If confidence in the ablity to detect is comparatively lower, then disclosure will happen sooner. Once use of the exploit has been detected, disclosure should follow immediately.

4.4 Is the vulnerability needed?

How badly do we need the intelligence we think we can get from exploiting the vulnerability? Are there other ways we can get it?

This factor is essentially the government's own estimation of the value of access to a device and the information it contains. If there are other vulnerabilities than can be exploited—or other methods entirely—with less cost or risk, then those other methods might be preferable.

Effect. The existence of other, less costly methods of obtaining the desired information will reduce the value of retaining this vulnerability and accelerate the timing of the disclosure. The availability of substitute methods depends on the nature of the information needed: concentrated info might be easier to

acquire with other means, whilst broad-based information, spread over a number of sources, might not be possible to acquire without the exploitation of the vulnerability.

4.5 Discovery by others

How likely is it that someone else will discover the vulnerability?

In a 2013 discussion about the government's approach to vulnerabilities [17], Hayden discussed the concept of 'Nobody But Us' (NOBUS) vulnerabilities, which the government believes others are unable to exploit:

> If there's a vulnerability here that weakens encryption but you still need four acres of Cray computers in the basement in order to work it you kind of think 'NOBUS' and that's a vulnerability we are not ethically or legally compelled to try to patch — it's one that ethically and legally we could try to exploit in order to keep Americans safe from others.

However, simultaneous discovery of a vulnerability may be relatively common. Schneier mentions several examples of simulatenous discovery [21]—including Heartbleed, which was discovered by both Google and Codenomicon. Studies of vulnerabilities in Microsoft software by Ozmnet [16] also suggest that simultaneous independent discovery is likely. More recently, a RAND report by Ablon and Bogart [1] followed a number of zero day exploits over time, and concluded that for a given collection of vulnerabilities, after one year 5.7% of them will have been discovered and disclosed by others. Another recent paper by Herr, Schneier, and Morris [10] studies a larger number of vulnerabilities and estimates that between 15% and 20% will be rediscovered within a year.

Different types of vulnerabilities probably experience different rates of independent discovery. If the government's ability to detect the use of known vulnerabilities by others is sufficient, they may be able to estimate how frequently simultaneous discovery occurs for different types of vulnerability.

Effect. If the vulnerability is likely to be discovered by others then it will accelerate disclosure. However, government confidence in a unique ability to discover or exploit some vulnerabilities will delay disclosure.

4.6 Can the vulnerability be used?

Could we utilize the vulnerability for a short period of time before we disclose it?

This can be interpreted in different ways. First, it may simply not be possible to develop an exploit for a particular vulnerability—not every bug found in software can be successfully exploited. Or, alternatively, this may refer to the time it takes to develop and utilize an exploit for this vulnerability. If exploit development takes a long time, it is more likely that either the information needed

will no longer be obtainable or no longer be of value, or that the vulnerability will be discovered and disclosed by another party. Another interpretation could be whether or not there is any benefit that can be gained by exploiting the vulnerability—perhaps the systems that could be accessed using the vulnerability have no intelligence or strategic value.

Effect. If the vulnerability cannot be utilized, then this will accelerate disclosure. If there is no benefit to be gained from retaining the vulnerability, then disclosing is the best option.

4.7 Can the vulnerability be patched?

Can the vulnerability be patched or otherwise mitigated?

There are a few reasons why it may not be possible to patch a vulnerability: some types of devices or software (for example, industrial control systems, SCADA systems, PLCs, or embedded devices) are rarely—or never—updated, and older devices or software may no longer be supported by the vendor. However, many of these vulnerabilities can be mitigated, if known, through additional security measures. There are cases when a vulnerability cannot be patched or mitigated. For example, old Android phones stop receiving security updates, and little can be done to mitigate this—other than switching to a newer device. In this case. disclosure of a vulnerability will not help users of the older devices (unless it encourages them to upgrade), but can help increase the security of newer devices if they share the same code.

Effect. If a vulnerability can truly never be patched or mitigated in any way, then it can lead to a considerable delay in disclosure—because doing otherwise will reveal the vulnerability to potential exploitation when the system can not be defended. However, this is unlikely. The speed at which a patch can be created and deployed may also have an effect on the disclosure timing. If patch creation and deployment is fast, then systems can be made secure more quickly if someone else discovers the vulnerability, which will delay disclosure.

5 Modelling the decision-making process

In considering whether to reveal the discovery of the vulnerability at any point in time, the government agency will consider the benefits and costs of the current situation—keeping the vulnerability undisclosed—and comparing them the the possible consequences after they have revealed the vulnerability to the public.

On one hand, retaining the vulnerability will allow the agency to access the information required for their purposes, and the longer the vulnerability persists—and the agency is able to exploit it undetected—the greater the potential accumulated benefit. On the other hand, if the vulnerability is not disclosed and remains unknown to the vendor and users, there is a chance that others will be able to exploit it, causing damage to the information assets the government

is charged to protect. This constitutes the expected loss to the government. The model we present here should be seen as a formalization of a thinking process; there is no hard data to populate the model, but it shows how the factors would be considered when making a decision.

In a general form, assuming continuous time, the benefits and costs the government will receive from not disclosing the vulnerability until a particular time, T, can be expressed as

$$B_T = \int_0^T \texttt{Benefit}(t)dt \qquad \text{and} \qquad C_T = \int_0^T \texttt{Cost}(t)dt,$$

which represent the total benefits and costs received from now until time T.

The government's aim is to find the best time to disclose the vulnerability,

$$\max_T V_T = B_T - C_T,$$

where V_t is the value to the government of disclosing at time t. This is shown in Figure 2, which shows the expected costs and benefits for disclosure at different times. The costs and benefits increase at different rates. The optimal time for disclosure maximizes the difference between costs and benefits. If the costs rise faster than the benefits, then the best action would be to disclose immediately.

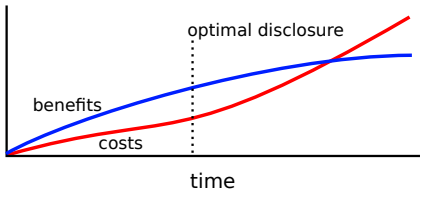

Figure 2. Total costs and benefits over time. The optimal timing for disclosure maximizes the difference between benefits and costs.

We relate the different factors discussed in the previous section to these benefits and costs. Any benefit of the vulnerability depends on the ability to use it (Can the vulnerability be used? — F_{use}). The benefits that the government expects to receive at time t depend on the extent of use of software (F_{extent}). It is not necessarily known in advance if there will be any use for the exploit at a particular time t; this depends on whether or not there is information that is needed at that time, or if there is a system to which the government requires access at that time. A greater extent of use increases the probability of a need for the exploit. The perceived value gained from exploiting the vulnerability will increase the benefits (Is the vulnerability needed? — F_{value}).

The expected cost from not disclosing the vulnerability will be rising with the extent of use (F_{extent}) and the possible harm that can result from its exploitation by others (F_{harm}). The ability to detect its exploitation by others reduces the

expected cost to the government (F_{detect}). Finally, the ability of others with high probability to exploit the vulnerability increases the cost of non-disclosure ($F_{discovery}$). The speed with which the patch can be developed and deployed (F_{patch}) reduces the expected cost of non-disclosure.

Immediate disclosure of the vulnerability upon discovery reduces the potential benefits to zero while minimizing the expected costs due to information assets damaged. However, this policy does not take into account the impact of the factors determining expected costs and benefits. Once such considerations are taken into account, the decision of when to disclose the vulnerability is equivalent to the solution of the problem to calculate the optimal timing for disclosure. In this context, the government is fully aware of both costs and benefits and their determinants and in effect decides when to exercise the 'option to disclose'. Intuitively, the decision will be such that at the time of the disclosure the marginal benefits from the retention vulnerability will be equal to the expected costs. Further delay in disclosing will result in the expected costs rising above the benefits. Although it is possible that such exact calculations cannot be made, the adoption of this equality as the organizing principle for the decision-making seems rational and more importantly, as it contains measurable quantities, it can be evaluated ex-post.

The analysis above is based on the assumption the the government is motivated 'equally/in a stable manner' by both benefits and costs. There are situations that call into question such a stable weighting. For example, in states of high alert, the benefits assume a far greater weight than the costs, compared to a normal situation where such immediacy of danger is not present. In this situation, the factors determining the benefits assume additional importance in the decision, resulting in delaying the disclosure of the vulnerability even though the expected costs are the same. This is because, in the eyes of the government, the value of the information obtainable through the use of the exploit is far higher.

5.1 Timing rules

We look at two different timing rules using, for simplicity, a discrete-time model. The first considers no delay, so disclosure happens at time $t = 0$. The second considers some delay, with disclosure at time $t = T$. For both timing rules, we can consider the benefits as immediate benefits (received at $t = 0$) plus discounted expected future benefits, with the same done for costs: $B = B_0 + B^e$ and $C = C_0 + C^e$.

For the first case, immediate disclosure, the immediate benefits, B_0, are zero because the vulnerability is disclosed at time $t = 0$, so the government has no chance to gain from its exploitation. Additionally, there will be no future benefits, so $B^e = 0$. For the second case, where disclosure is delayed, the government sees an immediate benefit B_0. The value of B_0 is determined by two factors, the value of the information and the ability to use the exploit to gain it, and can be written $B_0 = f(F_{value}, F_{use})$. The expected future benefits, B^e, are

$$B^e = \sum_{t=1}^{T} d_b^t \, \mathbb{E}\left[B_t\right],$$

where d_b is the discount factor applied to future benefits and $\mathbb{E}[B_t]$ is the expected value of benefits at time t. These expected benefits depend on all of the factors and evolve according to the time-evolution of the underlying factors. For example, were the extent of use to expand in the future, the expected benefits would increase because the likelihood of being able to access needed information using the exploit increases. If, in the future, the information that can be collected by exploiting the vulnerability is not needed, the value of future benefits will decline. The total benefits for retaining the vulnerability until a time T is

$$B = B_0 + \sum_{t=1}^{T} d_b^t \, \mathbb{E}[B_t] \, .$$

Next, we look at the costs of non-disclosure. Similarly to the benefits, these can be decomposed into two parts: the initial cost and the costs incurred during the time period before disclosure. For both immediate and delayed disclosure, the initial costs C_0 are zero, and for immediate disclosure, so are the expected future costs, C^e. In the case of delayed disclosure, the expected future cost C^e acquires a positive value and can be written as

$$C^e = \sum_{t=1}^{T} d_c^t \, \mathbb{E}[C_t]$$

where d_c is the discount factor applied to future costs, and $\mathbb{E}[C_t]$ is the expected value of costs at time t. The value of these expected costs will be influenced by the evolution over time of the factors mentioned above. These factors will affect both the probability of incurring the costs, which might be increasing with time, and the value of the losses which also might be functions of time.

Finally, the total costs until a time T can be written as

$$C = C_0 + \sum_{t=1}^{T} d_c^t \, \mathbb{E}[C_t] \, .$$

5.2 Optimal timing

The problem of the timing of disclosure can be reduced to the solution of

$$V_{T^*} = \max_{T} (B - C) \, ,$$

where V_{T^*} is the net benefit to the government when the vulnerability is disclosed at the optimal time, T^*. Substituting, we get

$$V_{T^*} = \max_{T} \left(B_0 + \sum_{t=1}^{T} d_b^t \, \mathbb{E}[B_t] - \left(C_0 + \sum_{t=1}^{T} d_c^t \, \mathbb{E}[C_t] \right) \right) .$$

Each element B_0, B_t, C_0, C_t, of the equation is a function of the different factors and, in the case of B_t and C_t, also of time. We consider B_0 to be influenced

primarily by the value of information needed, F_{value}, and the ability to use the vulnerability, F_{use}. At time $t = 0$, it is known which information is required and available through use of the exploit, and as such, its value can be determined. However, if it is not possible to use the exploit (F_{use}), then the value of B_0 is likely to be very low. Future expected benefits, B_t, are influenced by the same factors, but are also influenced by the extent of use F_{extent}. At some time in the future, it may be that there is information needed that is available through the use of the exploit. If the extent of use is larger, then it is more likely that such benefits will be available; if the extent is lower, it is less likely. For the initial costs, C_0, the value is always 0; none of the factors influence this. This is because costs accrue over time, and at $t = 0$, no time has passed. Expected future costs, C_t, are influenced by a host of factors. The extent of use, F_{extent}, will influence positively costs as a greater number of information assets are exposed to the possible exploit. These costs are increasing over time. As the risk and harm, F_{harm}, increases, so will the value of expected future costs. If the risk of discovery of the exploit by others increases $F_{discovery}$, this also increases the expected future costs, while the ability of the government to detect (F_{detect}) the use of the exploits by others will reduce such costs.

Table 1 shows how the different factors affect the benefits and costs, compared to immediate disclosure, and their influence on the timing of disclosure. This gives a general picture of how the factors affect timing. With a richer model of how the costs and benefits arise for each factor, it would be possible to have a deeper analysis of the timing problem.

Factor	F_{extent}	F_{harm}	F_{detect}	F_{value}	$F_{discovery}$	F_{use}	F_{patch}
Benefits	+			+		+	
Costs	+	+	-		+		-
Timing	-?	-	+	+	-	+	+

Table 1. Influence of factors on the costs and benefits, compared to immediate disclosure, and how they affect the timing of disclosure. While F_{extent} influences both benefits and costs, it will likely have a greater influence on costs, moving disclosure forward.

5.3 Making decisions

The model above shows how each of the factors can affect the timing of disclosure. This is useful for understanding, in a general sense, how the decision of when to disclose depends on the interactions of the different factors, but would not be useful for actually making such a decision. To make decisions with this type of model, it must first be parametrized: estimates for how the expected values of each factor change over time are needed.

Only the government knows how it estimates and weights the different factors, and making accurate estimations is probably extremely difficult. However, given that the decision-making process is supposed to be 'biased toward responsibly disclosing vulnerabilities', any estimations *should* err on the side of caution by overestimating costs and risks, and underestimating the values of benefits.

The same should be done for discount factors: by reducing d_b, the discount factor for benefits, compared to d_c, the discount factor for costs, future costs will outweigh potential future benefits, and move the timing decision forward.

Even if it is impossible to determine the exact time for optimal disclosure, having an estimate can still be useful. If retained vulnerabilities are periodically reviewed, the estimated optimal time of disclosure could be used to set an upper bound on the time before the next review. With conservative estimates for the factors, this would help ensure that retained vulnerabilities can be reconsidered (with updated information) in good time.

6 EternalBlue and WannaCry

Recent events have shown that the decisions the government makes about whether to disclose or retain vulnerabilities can have significant repercussions. The WannaCry malware, which severely affected businesses and hospitals around the world is an excellent example. The malware used a vulnerability from a NSA-developed exploit known as EternalBlue, which was leaked to the public by the ShadowBrokers on April 14, 2017.

The vulnerability used in the EternalBlue exploit would only have been considered under the VEP if it was discovered after the introduction of the VEP in 2010. According to the Washington Post [13], EternalBlue was used for 'more than 5 years', implying that it would have been considered under the VEP—for the following discussion, we will assume that this is the case.

In discussions about the VEP (for example, in [23]), there is a tendency to think of the VEP decision as a binary choice: either disclose or retain. We have argued that this should be viewed instead as a timing decision: not *if* a vulnerability should be disclosed, but *when*. When the EternalBlue exploit was leaked to the public in April, Microsoft had already created and released patches for the 0-day vulnerabilities in March—presumably after being informed by the NSA it they became aware of the ShadowBrokers leak. The initial decision here was to retain and use the vulnerability in EternalBlue, but to disclose it when it became clear it had leaked and could cause losses; it was a matter of timing.

While we do not know if the government makes decisions using the approach we described above, it can still be a useful tool for analysing the government's disclosure decisions. For the WannaCry/EternalBlue example, there different possible interpretations. The first case is that the decision was made using a correct model. This implies that the vulnerability was disclosed at the appropriate time, and that the benefits gained from the long-term retention of the vulnerability were valuable enough that they where not outweighed by the damages and costs that arose from the leaked vulnerability and resulting malware. The second case is that the timing of the disclosure was wrong because the model was missing a factor: the possibility of a vulnerability being leaked. From the Daniel blog post, we know that the risk of independent discovery is considered when making a decision, but it is unknown if this also includes the risk of leaks. If not, then the time of disclosure would have been after the optimal point. The final case

is where the timing of the disclosure was wrong because the model's parameters were incorrect. First are the extent of use and patching factors: even though the patch was released by Microsoft before the WannaCry malware, many computer systems were still vulnerable, either because the patch had not yet been applied or because they were running older versions of Windows that were out of support and so did not receive the patch. If the rate at which patches can be developed and applied is overestimated, or the number of systems running software that is no longer supported is underestimated in the model, then the potential costs will be underestimated resulting in a non-optimal, later time of disclosure. Incorrectly underestimating the probability of a leak (possibly included in the discovery factor) would also cause such a delay in disclosure.

Without knowing how much value the government gained from use of the exploit, a detailed understanding of the factors used when making a decision and how they are calculated and weighted, it is impossible to know which, if any, of these cases is true. However, WannaCry caused a lot of damage and could have caused a lot more, had it not been stopped. It is unlikely that this was anticipated and accepted, and therefore unlikely that the first case is true.

The remaining two cases suggest some possible improvements to the decision-making process. First, if the risk of vulnerabilities leaking is not included, it needs to be added. Second, a better understanding of how systems are patched over time may be needed when deciding when to disclose. Many older machines running out of date software are still used in critical processes; the costs of attacks on these machines must be considered. It may also be beneficial to disclose before these machines become out-of-support or to reduce potential costs by sponsoring the creation of patches for out-of-support software still widely in use when the vulnerability is finally disclosed.

7 Conclusions

Government disclosure of vulnerabilities is important, but so is the ability of the government to conduct intelligence, offensive national security, and law enforcement tasks. It would be a mistake to immediately disclose every vulnerability discovered, but it would also be a mistake to disclose none. Recommendations for and proposed legislation about the VEP include periodic reviews of any retained vulnerabilities, allowing them to be used for a time before disclosure.

We have presented a model that shows how the different factors used in the decision can be combined to determine the optimal time to disclose. Understanding how the different factors affect the timing allows the decisions about vulnerabilities made by the government to be better interpreted. We have looked at the case of the WannaCry malware, which used a leaked NSA zero day vulnerability. The vulnerability was disclosed to Microsoft before the malware was created, but before that remained undisclosed for 5 or more years.

It is likely that the disclosure came after the optimal time, as many systems remained unpatched and were vulnerable to WannaCry. The government could have underestimated or ignored the risk of the vulnerability leaking, or over-

estimated the speed with which systems could be patched. In any case, future decisions should include or improve the estimation of these factors in order to better determine the optimal time of disclosure.

References

[1] Lillian Ablon and Timothy Bogart. 'Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits'. In: (2017).

[2] Yolanta Beres, Jonathan Griffin and Simon Shiu. 'Security analytics: Analysis of security policies for vulnerability management'. In: Technical Report HPL-2008-121, HP Labs. 2008.

[3] Bill Budington and Andrew Crocker. *NSA's Failure to Report Shadow Broker Vulnerabilities Underscores Need for Oversight*. https://www.eff.org/deeplinks/2016/09/nsas-failure-report-shadow-broker-vulnerabilities-underscores-need-oversight. Sept. 2016.

[4] *Commercial and Government Information Technology and Industrial Control Product or System Vulnerabilities Equities Policy and Process*. https://www.eff.org/files/2015/09/04/document_71_-_vep_ocr.pdf.

[5] Michael Daniel. *Heartbleed: Understanding When We Disclose Cyber Vulnerabilities*. https://obamawhitehouse.archives.gov/blog/2014/04/28/heartbleed-understanding-when-we-disclose-cyber-vulnerabilities. Apr. 2014.

[6] Denelle Dixon-Thayer. *Improving Government Disclosure of Security Vulnerabilities*. https://blog.mozilla.org/netpolicy/2016/09/19/improving-government-disclosure-of-security-vulnerabilities/. Sept. 2016.

[7] Mailyn Fidler and Trey Herr. *PATCH: Debating Codification of the VEP*. https://lawfareblog.com/patch-debating-codification-vephttps://lawfareblog.com/patch-debating-codification-vep. May 2017.

[8] Andy Greenberg. *Shopping For Zero-Days: A Price List For Hackers' Secret Software Exploits*. https://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/. Mar. 2012.

[9] Jason Healey. 'The U.S. Government and Zero-Day Vulnerabilities: From Pre-Heartbleed to Shadow Brokers'. In: *Journal of International Affairs* (Nov. 2016).

[10] Trey Herr, Bruce Schneier and Christopher Morris. *Taking Stock: Estimating Vulnerability Rediscovery*. https://ssrn.com/abstract=2928758. Mar. 2017.

[11] Joseph Menn and John Walcott. *Exclusive: Probe of leaked U.S. NSA hacking tools examines operative's 'mistake'*. http://www.reuters.com/article/us-cyber-nsa-tools-idUSKCN11S2MF. Sept. 2016.

[12] Charlie Miller. 'The legitimate vulnerability market: Inside the secretive world of 0-day exploit sales'. In: *In Sixth Workshop on the Economics of Information Security*. 2007.

[13] Ellen Nakashima and Craig Timberg. *NSA officials worried about the day its potent hacking tool would get loose. Then it did.* https://www.washingtonpost.com/business/technology/nsa-officials-worried-about-the-day-its-potent-hacking-tool-would-get-loose-then-it-did/2017/05/16/50670b16-3978-11e7-a058-ddbb23c75d82_story.html. May 2017.

[14] *National Security Policy Directive 54.* https://fas.org/irp/offdocs/nspd/nspd-54.pdf.

[15] ODNI Public Affairs Office. *Statement on Bloomberg News story that NSA knew about the "Heartbleed bug" flaw and regularly used it to gather critical intelligence.* https://icontherecord.tumblr.com/post/82416436703/statement-on-bloomberg-news-story-that-nsa-knew. Apr. 2014.

[16] Andy Ozment. 'The likelihood of vulnerability rediscovery and the social utility of vulnerability hunting'. In: *Workshop on Economics and Information Security*. 2005.

[17] Andrea Peterson. *Why everyone is left less secure when the NSA doesn't help fix security flaws.* https://www.washingtonpost.com/news/the-switch/wp/2013/10/04/why-everyone-is-left-less-secure-when-the-nsa-doesnt-help-fix-security-flaws/. Oct. 2013.

[18] Michael Riley. *NSA Said to Have Used Heartbleed Bug, Exposing Consumers.* https://www.bloomberg.com/news/articles/2014-04-11/nsa-said-to-have-used-heartbleed-bug-exposing-consumers. Apr. 2014.

[19] David E. Sanger. *Obama Lets N.S.A. Exploit Some Internet Flaws, Officials Say.* https://www.nytimes.com/2014/04/13/us/politics/obama-lets-nsa-exploit-some-internet-flaws-officials-say.html?_r=1https://www.nytimes.com/2014/04/13/us/politics/obama-lets-nsa-exploit-some-internet-flaws-officials-say.html. Apr. 2014.

[20] Bruce Schneier. *Managed security monitoring: Closing the window of exposure.* http://www.keystoneisit.com/window.pdf. 2000.

[21] Bruce Schneier. *Simultaneous Discovery of Vulnerabilities.* https://www.schneier.com/blog/archives/2016/02/simultaneous_di.html. Feb. 2016.

[22] Bruce Schneier. *The Vulnerabilities Market and the Future of Security.* https://www.schneier.com/blog/archives/2012/06/the_vulnerabili.html. June 2012.

[23] Bruce Schneier. *WannaCry and Vulnerabilities.* https://www.schneier.com/blog/archives/2017/06/wannacry_and_vu.html. June 2017.

[24] Ari Schwartz and Rob Knake. *Government's Role in Vulnerability Disclosure.* http://www.belfercenter.org/publication/governments-role-vulnerability-disclosure-creating-permanent-and-accountable. June 2016.

[25] Zerodium. *How to sell your 0day exploit to ZERODIUM.* https://zerodium.com/program.html. Mar. 2017.