An Attacker's Day into Human Virology

Axelle Apvrille, Guillaume Lovet

March 2012

1 Introduction

The Anti-Virus industry has obviously taken much vocabulary from medicine. Viruses, infection, replication, anomaly, behaviour are typical words which have a related meaning in biology and computer science. Several researchers have already tried to draw a parallel between both worlds - humourously or not. Commonly, the human body is compared to a computer [Kor09], like in Science Fiction novels, but other comparisons exists: for instance, [Gla98] compares a computer to a human cell. Anti-virus practices such as anomaly detection have also been compared to or inspired by biology [KSA+95, GTA06].

In this paper, we wish to focus on *attack scenarios* of biological and computer viruses. How much did biological viruses actually invent and how many 0-days could they be attributed? In this paper, after some reminders on medical virology (section 2) and the immune system (section 3), we speak for the voiceless, biological viruses, and give them credit for their inventions (section 4). We also compare cures for humans or computers (section 6).

On the other hand, a few devilish ideas have arisen from computer science (section 5). They fortunately do not apply to human cases, or should we say "they do not apply yet" when considering the advances in cybernetics? (see section 7).

2 Medical virology background

As defined by [Wike], viruses are *"small infectious agents that can replicate only inside the cells of other organisms"*. They are at the frontier between living and non-living organisms [Hun93]. They consist of two or three parts: genes made from either RNA or sometimes DNA (long molecules that carry genetic information or instructions), a protein coat protecting these genes and optionally, an envelope of fat surrounding them. Many viruses also develop spikes on their envelope, which help them attach to specific cell surfaces.

Viruses should not be confused with bacteria [Wis]. Bacteria are complete one-celled living organism, with a complete set of genetic codes. They are consequently much bigger than viruses which commonly measure less than 250 nm, i.e 250.10^{-9}m. Also, bacteria are self-replicating units: they do not need another host to replicate as viruses do.

Once a virus enters the human body, it looks for an acceptable host cell. When it has spotted a suitable target, it clamps onto the cell and penetrates the cell using a spike or a chemical coating. Then, there are two different ways a virus replicates, this is the lytic or the lysogenic cycle (see Figure 3). In the lysogenic cycle, the DNA of the virus

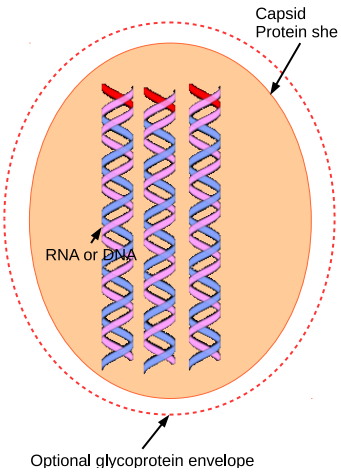

Figure 1: Structure of a virus

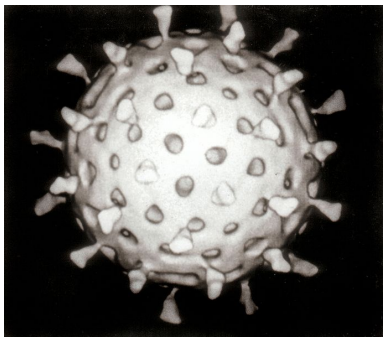

Figure 2: Computer assisted reconstruction of a rotavirus particle - Graham Colm

is combined to the host cell and the host cell replicates with the infected code. In the lytic cycle, the virus's DNA or RNA is injected in the host cell and makes numerous copies of itself. In the end, it bursts the cell's membrane and releases new viruses.

3 The immune system

To defend against viruses, the human body uses the immune system [Wikb, Nob].

The *innate immune system* provides a generic way to fight any kind of viruses, while the *adaptive immune system* specifically detects the intruder and then massively develops the production of specialized lymphocytes to fight back the virus.

Besides the immune system, note the skin and mucuous membranes themselves consist in a first shield against viruses.

3.1 The innate immune system

Usually, unknown entities are first detected by *complement proteins* that flow in the blood. Those complement proteins alert phagocytes to come and eat up (phagocytose) the intruder.

There are three main types of phagocytes: *granulocytes*, the faster to react but with a smaller appetite, *macrophages* (big appetite but slow to react) and *dendritic cells* (similar to macrophages, but in contact with the external environment such as in the skin, nose or lungs).

Additionally, macrophages release special proteins - cytokines - that activate Natural Killer lymphocytes [Wikc]. Those cells kill infected or damaged cells they encounter by releasing cytotoxines that create pores in the infected cell's plasma membrane. This weakens the cell so that other toxins released by the NK cells finally kill the infected cell.

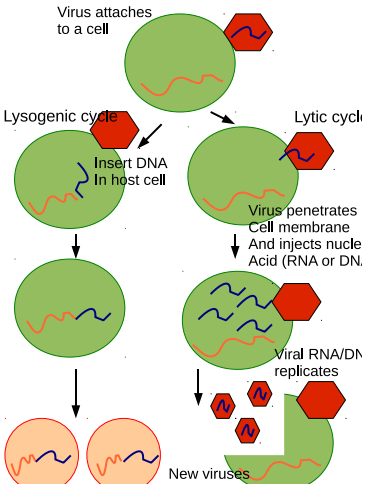

Figure 3: The lytic and the lysogenic replication process

3.2 The adaptive immune system

The specific response of the adaptive immune system are handled by B and T lymphocytes.

B and T lymphocytes have specialized receptors matching only a given type of intruder. This is why this part is called the *adaptive* immune system. Fortunately, the body contains many different B and T cells, hence protecting against many different intrusions.

After a phagocyte has eaten up a virus, it reports to the nearest lymph node to provide information about the intruder. Lymph nodes act like garrisons of immune cells and trap foreign cells. They are found all over the body, especially under the armpit and in the stomach. A special T cell called a 'helper T cell', handles the information, starts to divide and produce cytokines that activate other lymphocytes.

Helper T cells can be seen as managers: they do not act against the intruder themselves but give orders to others lymphocytes to carry out the task. The lymphocytes which actually do the dark work are killer T cells and B cells.

Killer T cells are meant to clean up the body from infected or damaged cells. They act like natural killer cells, except they are dedicated to a given type of virus.

B cells flowing in the body also have the capability to detect cells infected by a given type of virus. When this happens and they are activated by helper T cells, they divide into plasma cells which secrete plenty of antibodies, i.e specific proteins able to handle that particular infection. These antibodies mark infected cells so that they are more easily spotted by phagocytes. Sometimes, antibodies are able to directly incapacitate the virus. Killing infected cells, and incidentally the virus it contains, controls the replication of the virus, and in the end, eliminates it.

The adaptive immune system also implements a memory mechanism. When B or T cells divide, some of their offsprings become B or T memory cells (see Figure 4). Those cells remember the virus they met (months or years after) so that the immune system can react faster next time. Vaccines use this functionality, presenting disabled or weakened viruses to the body, so that the memory cells keep in mind the attack and

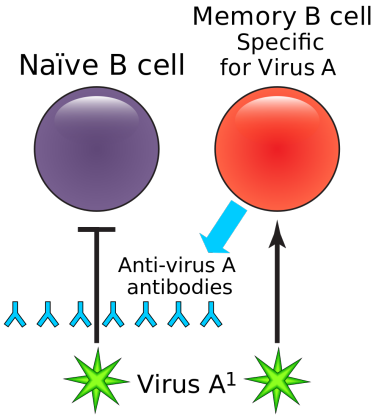

Figure 4: A naive B cells hasn't ever met an antigen. Memory B cells are created after the immune system has fought against a given intruder. Image courtesy of Wikipedia

react faster to real invasions.

4 Biology's inventions

4.1 Brute-force attacks: outnumbering defenses

One of the typical attack strategies of biological viruses consists in outnumbering defenses of the immune system. For example, at some stages of the disease, people infected with AIDS can have their blood carry over 1 million of HIV virus per ml [PGF93]. Also, whenever patients infected with Influenza (flu) speak or breathe, they release aerosol particles. A mere 0.1 μl of aerosol particles contains more than 100 virus particles [Gur].

Comparatively, modern computer viruses don't have the same urge to massively infect the host they are on. They usually only infect specific files on the host, such as files of the Windows (`c:\windows` or Program Files `c:\Program Files` (e.g Win32/Xpaj [Mic10]). There are even cases where computer viruses only infect once and make sure not to re-infect the host, for instance to prevent bugs. This is what Symbian malware Yxes implements as a basic semaphore. When the malware is launched, it tries to open a given global semaphore. If the semaphore exists, this means another instance of the malware is running, so the program exits. Otherwise, it creates the semaphore [Apv10].

Today, malware authors are driven by money [Lov06]. Massively infecting a given host doesn't bring in more money (it's infected that's all that counts). Rather, they are interested in propagating the malware to other hosts, and hence new victims. Computer worms are excellent at this. In four days, Conficker infected 6.5 million new hosts, raising the amount of infected computer to 8,976,038 hosts [Koi09]. Before that, the Slammer worm infected more than 90 percent of vulnerable host in 10 minutes [Dav03]. Computer viruses use several different ways to propagate. [Sym10] list the 10 top most propagation mechanisms, such as email attachments, URIs in Instant Messages, and there are yet others methods, such a P2P (used by Conficker) or drive-

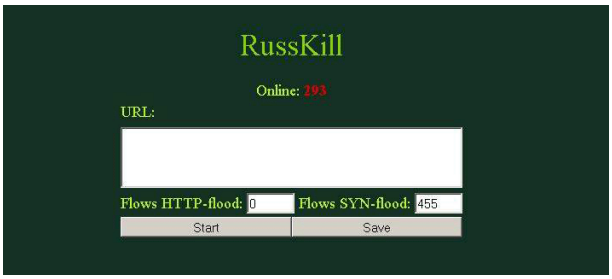

Figure 5: RussKill's control panel - image courtesy of [NoV10]

by-downloads and phishing schemes: this is how the ZeuS trojan built a botnet of an estimated 3.6 million hosts in the USA alone [Oll10].

Most mobile phone viruses simply sit and wait on a highly visited download web site and infect each unlucky victim who downloads them. For example, Android/DrdDream infected several applications of the Android Market (Super Guitar Solo, Photo Editor etc) [For11] and propagated to over 250,000 mobile phones by just being available on the famous marketplace. In that case, the Android Market acted like the waiting room of a busy medical doctor.

Actually, biology's outnumbering strategy is more comparable to DDoS bots. Russkill, for example, is able to command to all of its bots to perform an HTTP flood or a SYN flood on a list of hosts [NoV10]. It put the KrebsOnSecurity.com down in November 2011 [Mau11].

4.2 Polymorphism: mutating to evade defense

In biology, the process of replication consists in duplicating its genetic information, the DNA. Fortunately for viruses (and unfortunately for patients), this process is error prone, so that copy errors occur once in a while (by default) or more often if the virus embeds shrewder techniques. For example, the Influenza virus maliciously omits the replication error checking protein. Consequently, the replication of Influenza contains more errors - read variants for AV analysts - than usual: on average, each copy of Influenza contains a random mutation [Hua09].

The HIV virus shows the same error-prone replication property. [SMD+09] estimates there is one substitution per genome per replication round. This polymorphism makes it difficult for killer T cells to recognize the virus and slows down the response of the immune system. HIV's polymorphism also accounts for the difficulty to design a good vaccine against it.

In computer science, polymorphic viruses also are the nightmare of analysts, because each replicated sample is significantly different from its ancestor. Thus, hash-based signatures to detect the virus are close to useless. Analysts typically have to reverse engineer the virus and design a generic signature based on invariants or behaviour. The famous Conficker virus is polymorphic, so are several others: W32/Xpaj [Zay09], Sality, Mabezat, Koobface...

Besides polymorphism, biological viruses are also prone to mixing. [Hua09] gives the example of a person simultaneously catching different types of flus. Then, the replication process produces yet another variant of flu, combining various parts among different types. The case also occurs for HIV [FBRSB10]: when a given cell is infected

by two different subtypes of HIV viruses, they share genetic material and produce a third hybrid sort. For computers, mixing can also occur when, for instance, a worm is infected by a file infector. Then, the worm propagates hybrid infected files. [Man09] precisely acknowledges such a case where a computer initially infected with MyDoom (worm) also got infected with Virut (file infector). Virut infected the MyDoom sample, and actually, a modified version of MyDoom replicated and sent itself to other victims, propagating both MyDoom and Virut in one sample.

4.3 Attacking the AV engine

HIV is a well known virus which targets what would correspond to the AV engine on computers. Indeed, it particularly infects the immune system itself, reproducing in helper T cells, macrophages and dendritic cells. The fact helper T cells are infected - and then die - lowers the numbers of sane cells, disorganizes the action of killer T cells and B cells, and quickly leads to immuno deficient. Note that HIV is not the only virus to attack the immune system. The flavivirus, responsible for the yellow fever, is known to replicate in lymph nodes and dendritic cells [Wikf].

This strategy also shows with computer viruses such as W32/Sality which terminates running anti-virus programs and also sets the virus as an authorized application to bypass Microsoft's firewall (HKLM \CurrentControlSet \Services \SharedAccess \Parameters \FirewallPolicy \Authorized Application List).

4.4 Finding vulnerable hosts

At first sight, it seems that biological viruses attack at random cells they encounter. As usual, reality is however more complex, as some viruses attack some particular cells and not others. For example, the Rotavirus (which causes severe diarrhoea) targets cells that line the small intestine, the Poliovirus (poliomyelitis) targets preferentially motor neurons, the Rhinovirus (causing up to 50% of common colds we are affected with every year) attaches to cells carrying special receptors, which are located on the lining of the nasopharynx etc.

Computer viruses show both behaviours: some attack at random any host and try to perform their malicious actions. This succeeds only if the host is vulnerable to this type of attack. For example, W32/Expiro collects credentials stored in FileZilla, Internet Explorer and Windows Protected Storage. If the end-user does not store any credentials in those software, the malware is ineffective: the attack is "blind", the malware just statistically hopes to retrieve sensitive information from some end-users.

Others do test the host is vulnerable to a given hack before infected it. For instance, iPhone's worm Eeki (aka Ikee) [For09] scans other iPhones and tests whether their root password is still set to the default value or not. If this is the case, the target is vulnerable and the worm propagates to it.

4.5 Sleeping beauty

Most biological viruses are not effective straight away. This is either because they haven't replicated enough yet and are not numerous enough, and/or because they do not immediately start to replicate. The Rhinovirus - though very contagious - only starts replicating 8 to 12 hours after the initial infection. And as a matter of fact, most diseases show what medecine calls an incubation period (i.e the time between infection

and first symptoms): for chicken pox, it is around 2 weeks, for the flu, 2 to 3 days, measles 6 to 19 days, ebola 2 to 21 days, rabies 2 to 12 weeks etc.

The same strategy occurs in what AV analysts call time bombs (viruses meant to release their malicious payload at a given time). In 1991, the Michelangelo virus for instance, was set to deploy on Michelangelo's birthday, March 6th. Some other viruses are also known to significantly change their behaviour at a given date. For example, the CodeRed worm (2001) tries to infected new hosts between the 1st and 19th of each month, while it only conducts a denial of service attack against the US whitehouse website (www.whitehouse.gov) from the 20th to the end of the month [PM01].

Currently, time bombs are not very popular and they are seldom seen in the wild. This is perhaps because the vast majority of virus writers now (nearly) only focus on business and money, and all dates are as good for that matter (apart from viruses which target e-commerce users: they are usually more profitable close to Christmas or other religious events). Conficker did have a date trigger (April 1st, 2009), but, basically, it just consisted in a new update of the virus, not a specific new behaviour for that date [Mik09].

4.6 Remaining infected

Some biological viruses are smart enough to develop a strategy to keep their host infected, or, if necessary, automatically re-infect it.

In the case of HIV, the strategy relies on infecting long-life cells (ensures a copy of the virus always exists) and making sure they won't be detected. T memory cells are the perfect target for this operation. Normally, memory cells are meant to keep track of previous infections over years and help the immune system respond faster to infections it has already encountered. However, in the case of AIDS, memory cells themselves are infected by the DNA of HIV. This is how HIV makes sure there are fresh copies of itself even after years of treatment. Furthermore, as the DNA of those memory cells are not used to create corresponding proteins, the cells do not produce any viral antigen and consequently, the immune system does not detect the cell is infected. Those infected cells can keep on infecting other cells in the immune system without being detected.

Keeping the host infected is another high concern of modern computer viruses. For example, bots like Zeus [BOB+10] are frequently updated by bot masters with newer versions of viruses, so that the infected host remains controllable by the cyber-criminals. Other malware, like TDL4, permanently infect a host by infected the Master Boot Record [Giu11].

5 Computer science inventions

Fortunately, biological viruses are not advanced enough to implement packing, encryption, virtual machine detection or anti-debugging tricks. Genetists should probably pray that future biological viruses won't evolve from the genomes of cyber-criminals, or medecine will even have more difficulties in identifying and stopping viruses.

We are quite lucky that biological viruses do not intentionally play a game of cat and mouse with medical doctors, whereas computer viruses are often seen to specifically redirect AV web sites to other URLS (e.g W32/DNSChanger), detect reverse engineering tools installed by AV analysts, detect execution inside virtual machines, detect they are being run within a debugger etc. [Fer08] wrote several articles on Anti-Debugging tricks found in malware. For instance, he describes a technique where the

malware detects different behaviours when calling the CloseHandle function: if a debugger is present, an EXCEPTION_HANDLE_NOT_CLOSABLE is raised, and can be used by malware to detect the presence of a debugger.

Just imagine what this would translate to: sick people being redirected to a hair stylist instead of a medical doctor for treatment, viruses saying hello when you watch them with a microscope, viruses refusing to replicate under culture etc.

The relative simplicity of biological viruses compared to computer ones is perhaps due to their code length. For example, [Hua09] has computed (from scientific sources) that the DNA of Influenza fits in approximately 23,000 bits, ie only 22KB. Although some computer viruses are that small, their average size is rather 10 tor 100 times bigger. This indeed leaves developers more space to implement advanced tricks.

6 Cures

It is difficult to tell which is the more efficient against viruses between an AV engine and the body's immune system. Some would probably argue that it depends on which AV engine or which body one compares... Nevertheless, the following subsections try to detail the various defense mechanisms both systems use, with their pros and cons.

6.1 Shoot everything suspect

The immune system undoubtely uses drastic measures to counter viruses which try to invade it. One should indeed notice that it kills all infected cells it detects. It does not directly kill the virus by itself but controls its replication and, in the end, eventually manages to eradicate it.

Unfortunately for AV analysts - but fortunately for business - the AV industry cannot seriously promote burning all infected devices to get rid of a virus: there would soon be no computer left (apart from OpenSolaris boxes - private joke). As much as possible, AV vendors provide cleaning engines which try and remove the virus from an infected host. Those could be compared in medecine to post-exposure treatments used against Rabies (PEP) as they are administered after infection. Those cleansing engines are not always easy to design, and not always entirely efficient. For instance, in cases of password stealing or file deletion, they are unable to recover the damage, like a disease may leave scars on the victim. The recent Conficker worm is a good illustration to the difficulty of erasing worms. It infects the hosts at several levels, hides on the system with different names, different patterns each time so that cleaners can easily miss some instances. In practice, many anti-virus analysts would probably recommend a full reinstallation of the infected host, from scratch (without recovering potentially infected mails, bookmarks or settings).

6.2 Generic signatures

The human body is fortunately quite efficient in design at recognizing new viruses it does not know about initially. Basically, this is because the immune system does not need to analyze the entire DNA of a virus to detect it.

Basically, there is genericity at the following levels:

- the body is automatically capable of detecting self cells from pathogens. For computers, detecting an executable is a virus is much more difficult. We are not

natively able to tell it is a virus without analyzing it. In biology, the body uses Major Histocompatibility Complex (MHC) (also known as the Human Leukocyte Antigen (HLA)).

- each cell (even specialized lymphocytes of the adaptive immune system) is able to bind to a few different pathogens. [Hof00].
- HIV vaccines try to focus on detecting non-variable elements, such as the Pol or Gag genes of the virus. Not the gp120 protein, because it varies too much. Not so successful yet.

The AV engine is probably less efficient than the body regarding generic signatures. There is only 1 % of generic signatures: approximately 10 million signatures, and only 100 000 generic ones, but of course, each generic signature hits far more often than any other signature.

6.3 AV Analysts at work

There are cases where the body encounters a really new viruses for which he does not have any prior immunization. There are a few chances the virus shall be blocked by the skin/mucuous membranes or the innate immune system, but of course, once in a while it might happen nobody manages to block it. In that case, the body's adaptive immune system is in charge and must block the intruder.

As each lymphocytes of the adaptive immune system only carry one type of receptors, each specific lymphocyte is (basically) only capable of stopping one type of antigen. Fortunately, the body carries 10^8 different type of lymphocytes, hence providing detection for 10^8 different antigens. Even if this is very much , it is still insufficient to block *any* kind of biological virus, as there is an estimated 10^{16} possibilities. But as some of the former 10^8 lymphocytes die, new ones are created and the creation process ensures that the new lymphocytes recognize some other pathogens. So, that after a few days or weeks, the 10^{16} possibilities have all been checked out.

The analogy with an AV team is stricking. If a virus is not caught by the 10 million signature that are already designed, a sample of the virus ends up in the AV queue. There, if possible the sample is processed automatically. If not, the sample must be processed manually. It gets in the mailing box of analysts, where an alarm is triggered to tell us we must hurry and tackle a new issue. The AV analyst replicates the virus in an isolated computer of the AV lab, analyzes what it does, reverse engineers it and, from this analysis, crafts a signature to detect the virus. Then, the signature is pushed onto all AV engines.

6.4 Prevention

Another way to tackle viruses consists in focusing on prevention. On this matter, medecine is perhaps more successful than the AV industry with the development of vaccines. There are vaccines for poliomyelitis, measles, mumps, rotavirus, rabies, yellow fever, chicken pox, HPV, rubella, influenza... Many of those are enforced by governments and have widely contributed to reducing the spread of those disease.

The computer world does not use vaccines. Actually, in medecine, vaccines are often built from real but weakened or disabled viruses. Today, doing the same against computers would probably be considered as unethical, mainly because of the risk of really getting infected and spreading the virus. As a matter of fact, this argument is

also given by detractors of medical vaccines, because there is sometimes a risk the weakened/disabled viruses reverts to an active form.

Health organizations also publish numerous recommendations trying to educate people so as to limit risks of getting infected. Vast campaigns have explained how AIDS would transmit, including practical recommendations. More recently, governments have focused on prevention of Influenza. In several countries, people have been wearing masks, and can't go to the toilets without reading a sign about how to wash hands. Computer users too have been warned numerous times not to answer spam, not to surf on unknown websites, keep their operating system up-to-date etc [Apv09]. Yet, it seems that both medical and computer education are insufficient to contain viruses, because 1/ it is difficult to change habits of people and 2/ recommendations are not always easily understood (for e.g what is an unknown or suspicious website ?).

7 Convergence and Keys to Future Threats

Along our biological vs computer virus comparison, we have mostly focused on functional aspects, so far - Comparing how viruses attack their target, and how the latter reacts. At this point, we may therefore go beyond function, and extend the comparison to essence and purpose, so as to open questions on convergence of biological and computer viruses... And the future threats such may yield.

7.1 Essence

As we pointed in the medical background section, a biological virus is essentially a strand of DNA (or RNA). Now, a strand of DNA is a sequence of nucleobases (A, G, C, and T), that codes behavior of the virus within its host system, when the infected cell produces the proteins that the sequence describes. In a nutshell: a biological virus is information that codes for behavior in a host system.

On the other hand, a computer viruses is a sequence of bits, that codes for behavior of the virus, when the infected OS runs the instructions that the sequence describes. In a nutshell: a computer virus is information that codes for behavior in a host system.

In essence, they are therefore the same: information that codes for parasitic behavior (replicative, and optionally, deleterious for the host). Seen from that angle, [Coh87] was particularly pertinent when he coined the computer virus term in the 80s .

7.2 Purpose

If back then, computer viruses had little more purpose than replicating, and later, impairing or destroying their hosts, it has dramatically changed recently. As evoked earlier, today, computer viruses' main purpose is making their masters (i.e. cybercriminals distributing them) rich [Lov06]; more rarely, the purpose may be political espionage (e.g.: Ghostnet), industrial espionage (e.g. Aurora), or physical destruction of a strategic target (e.g.: Stuxnet).

What about biological viruses? Framing the purpose of an entity that sits at the frontier of the living and the non-living is, to say the least, tricky, and quickly leads to endless philosophical, and possibly religious debates. Therefore, we will only point that while computer viruses, in their final state, are made of coding information (see above) that is designed by a conscious intelligence, biological viruses are the fruit of

random mutations, the most profitable ones in terms of survival and replication being selected by environmental pressure.

Now, would it be possible to witness, in the future, exactly the opposite? That is to say, biological viruses designed for a specific purpose, and computer viruses stemming from random mutations.

7.2.1 Designed Biological Viruses

Literature and pop culture are full of stories of man-engineered viruses used as biological weapons (e.g.: The infamous St Mary virus in "V for Vendetta") - not to mention the obligatory conspiracy theories, arguing that some of the deadliest viruses today were man made; those range from completely fanciful (e.g. AIDS, see [Wika]) to supported by a fistful of scientists (e.g. SARS, see [Wikd]).

However, no seriously documented case of a synthetic virus used as a biological weapon exists to our knowledge. Nonetheless, it would be technologically possible: in 2002, scientists successfully synthesized the polio virus [CPW02] as a PoC, and more recently, a "custom" version of SARS built from scratch, in order to "help explain how the SARS evolved" [Wir08]. Fortunately for us, it seems that using a virus as a military purposed biological weapon has a number of disadvantages over resorting to bacteria such as Bacillus Anthracis (a.k.a Anthrax), one being that its spreading is much harder to control, and could very well backfire at the attacking army. Beyond this, biological weaponry has been banned under the Geneva Protocol of 1925, a ban extended by the 1972 Biological and Toxin Weapons Convention (BWC). Of course, none of this prevents us from seeing bioterrorist attacks based on designed viruses one day.

7.2.2 Darwinian Computer Viruses

While unseen in the wild so far, computer viruses evolving along the Darwinian rules of replication / mutation / selection have, as a matter of fact, been created by researchers, as proof of concept pieces [NMSF09]. Unsurprisingly, they heavily resort to genetic algorithms. However, such viruses, even if introduced in the wild some day by cybercriminals (perhaps to implement some form of behavioral polymorphism, in order to fool behavior-based detection engines...), carry a significant dose of intelligent design to begin with. Could such a virus be actually spontaneous (like biological viruses are thought to be)?

In the science-fiction masterpiece "Ghost in the Shell", the flow of digital information running throughout the World has become so dense, that it accidentally gave birth to a sentient form of life. This extends far beyond the simple "creature turning against its creator" ubiquitous scenario (from Selley's Frankenstein to Terminator's Skynet). Without going as far as spontaneous sentient life creation, would it at least be possible that a computer virus appears out of the data flowing on the wires around the World?

No documented case of such exists, yet the question is not so incongruous. Security researchers, more than anyone, know that software is full of bugs, and that presented with particular inputs, the execution flow may be diverted to arbitrary or unexpected portions of the memory. What if that portion contains data that, accidentally, forms the code of a simple virus? Unlikely? Yes, and even more so if we expect that virus to have capacities to evolve.

But impossible? No. Viruses can be as small as tens of bytes. As a matter of fact, in "A short course on computer viruses" [Coh94], Dr Fred Cohen states that "The record for the smallest virus is a Unix sh command script [...] that takes only about

8 characters", and that "If you want a virus that evolves, replicates, and does damage, you need about 5 or 5 lines". Now, according to IBM [Cam07], back in 2007, "15 petabytes of new information was generated daily, the codified information base of the world doubling every 15 hours". The probability that among this data, consecutive bytes form the code of a virus, and that at some point, the execution flow of a CPU gets directed to those bytes is certainly not null. Evaluating such a probability precisely is beyond the scope of this paper, and left as an exercise to the reader.

7.3 Convergence

We have noted above that although generally having a different origin, computer and biological viruses were the same, in essence: Information coding for a parasitic behavior inside a host system. Of course, that information is materialized differently - by molecules in the biological case, and by electro-magnetic properties in the computer case. Thus, the question "Will there be cases of convergence in both types of viruses one day (i.e. a biological virus passing in the computer realm or vice-versa)?" may sound foolish at first. We will see that it is perhaps not that devoid of sense.

7.3.1 Blurring the frontier

For starters, there are interfaces between the computer and the biological worlds, that, if not allowing to actually cross the frontier, at least blurs it. Cybernetic prosthesis are a good example: Some people have various electronic devices embedded in their bodies, ranging from pacemakers to deep-brain stimulators, cochlear implants, etc... And it will likely increase in variety and adoption rate as cybernetics get better. Most of these device embed, or will embed, chips with operating systems. In other words: computers. As soon as those computers can communicate with external devices (which in most cases is, at some point, necessary), they become theoretically vulnerable to computer viruses.

As a matter of fact, in 2010, a scientist at the University of Reading, as a proof of concept, infected the RFID chip he had implanted under his skin (and that he would commonly use to authenticate wirelessly in various places at the university) with a basic RFID virus, that infected the reader systems [BBC10]. Granted, RFID viruses have been known for years (he used a simple self-referencing SQL injection-based virus), but the stunt does prove a point: A virus can hop from a "new man" to a computer (and theoretically, vice-versa).

Here, it is not the computer virus that evolved to pass into the biological world, but rather the biological world itself -or at least, the definition of a living organism within it- that evolved, to permit that.

7.3.2 Bridges across the frontier

We have mentioned earlier that as early as in 2002, scientists were able to synthesize the polio virus. Biotechnology has advanced since then, and in 2010, genome pioneer Craig Venter managed to synthesize a bacteria [New10] , by implanting a full synthetic genome in a cell. Beyond that, every day, living organisms are genetically modified, synthetic genes being implanted, for industrial and medical applications.

Now, where is stored the information that code for all the synthetic DNA generated in applications above? Right, on computers... Seeing that the infamous Stuxnet virus, in 2010, was able to creep through a uranium enrichment plant, seize control of its PLC,

and destroy its centrifuging gear, one could reasonably think that a virus infecting the computers sporting DNA databases is not outside the realm of possibility. From there, the virus could very well inject a parasitic replicative sequence in the genes being synthesized, and see it grown in lab (or worse, at industrial scale). Thereby hoping from the computer into the biological realm.

Conversely, software used when sequencing DNA of a living organism, and databases storing bits that code for that sequence, are probably not absent of vulnerabilities. Whether it is possible to craft a virus with malicious DNA sequences that could, once transcribed into bits, exploit those vulnerabilities, is an interesting question. Let's imagine a software piece that stores genomes of new bacteria and viruses as they are analysed in lab, classified by type, and expecting a certain length of DNA for each type. With no input validation... A quite exotic way to take control of the back-end database, possibly containing extremely sensible information. Such as the smallpox genome (actual samples of smallpox being held at only 2 locations in the World: The Russian State Centre for Research on Virology and Biotechnology, Koltsovo, Russia, and Centers for Disease Control and Prevention, Atlanta, Georgia, USA [Sec03]).

8 Conclusion

Along our attacker's day into human virology, we could compare the merits of computer and biological viruses in terms of arsenal, strategies, and attacking power. We noted in the process that some inventions in the computer immune system have been inspired by biology, and we've drawn parallels between cures for infected computers and infected humans. Beyond functionality, we went one step further and discussed the essence, origin, and purpose of viruses in both worlds, which lead us to examine the question: will a virus be crossing the frontier between the computer realm and the biological realm one day? Providing a clear-cut answer was not possible, but we advocated that it is theoretically very possible. And yes, if we wanted to provide a proof-of-concept demo of such a virus, it would not have been silently included in the paper's PDF, and it wouldn't be transmitted through a simple laptop; so you may now remove that chemical protection gloves and that gas mask, it will be somewhat easier to read through the appendix section!

9 Appendix: Vocabulary conversion table

Biological world	Computer world
Virus	Computer Virus or Malware
DNA	Code of the virus
RNA	Virus loaded into the memory [Hua09]
Genes	Assembly functions
Protein coat protecting the virus	Packer protecting the malware
Virus seeking for an adequate host	Malware checking the right software or version is installed
Virus using a spike or chemical coating to penetrate the host cell	Malware using code injections or other tricks to infect the host
Human body	Computer
Cell	File
Virus replication, cell division	Malware execution, code injection
Long range communication (air, water, food...)	Internet, Mail
Short range communication (cough, fluids...)	Bluetooth, USB peripherals, network shares
Medical Doctor or Genetist	Anti Virus Analyst
Medical Research	Security Research
Waiting room	Drive by download website
External barriers	Firewall
Innate immune system	Generic signatures of an Anti Virus engine
Adaptive immune system	Specific signatures of an AV engine
Complement system	Intrusion detection system
Lymphocyte antigen receptors	Anti Virus detection engine
Killer T cells, natural killers and phagocytes	Anti Virus cleaning engine
Helper T cell	Manager
Antibodies	?
Cytokines	Alarm?
Infected cells marked by antibodies	Quarantined infected files
Memory cells	Virus database
First discovery of a virus	security advisory
Health report (e.g from the OMS)	security alert bulletin

Table 1: Vocabulary conversion table

References

[Apv09] Axelle Apvrille. Keep your phone healthy: H1N1 vs. SymbOS/Yxes, 2009. http://blog.fortinet.com/keep-your-phone-healthy-h1n1-vs-symbosyxes/.

[Apv10] Axelle Apvrille. Symbian Worm Yxes: Towards Mobile Botnets? In *Proceedings of the 19th EICAR Annual Conference*, pages 31–54, Paris, France, May 2010.

[BBC10] First human infected with computer virus, May 2010. http://www.bbc.co.uk/news/10158517.

[BOB+10] H Binsalleeh, T Ormerod, A Boukhtouta, P Sinha, A Youssef, M Debbabi, and L Wang. On the analysis of the zeus bot-

net crimeware toolkit. *Communications*, pages 31–38, 2010. http://www.ncfta.ca/papers/On_the_Analysis_of_the_Zeus_Botnet_Crimeware.pdf.

[Cam07] Don Campbell. Business Analytics Today and Tomorrow, 2007.

[Coh87] F. Cohen. Computer viruses: theory and experiments. *Comput. Secur.*, 6(1):22–35, February 1987. http://all.net/books/virus/index.html.

[Coh94] Dr Frederick Cohen. *A Short Course on Computer Viruses*. John Wiley & Sons, Inc., New York, NY, USA, 2nd edition, 1994.

[CPW02] Jeronimo Cello, Aniko V. Paul, and Eckard Wimmer. Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template, August 2002.

[Dav03] David Moore and Vern Paxson and Stefan Savage and Colleen Shannon and Stuart Staniford and Nicholas Weaver. The Spread of the Sapphire/Slammer Worm, 2003. http://www.caida.org/publications/papers/2003/sapphire/sapphire.html.

[FBRSB10] Emanuele Fanales-Belasio, Mariangela Raimondo, Barbara Suligoi, and Stefano Butto. HIV virology and pathogenetic mechanisms of infection: a brief overview. *Ann Ist Super Sanita*, 46(1):5–14, 2010.

[Fer08] Peter Ferrie. Anti-Unpacker Tricks Part One. *Virus Bulletin Magazine*, pages 4–8, December 2008.

[For09] iPhoneOS/Eeki.B!worm, 2009. http://www.fortiguard.com/av/VID1138180.

[For11] Android/DrdDream.A!tr, 2011. http://www.fortiguard.com/av/VID2523021.

[Giu11] Marco Giuliani. x64 TDL3 rootkit - follow up, August 2011. http://www.prevx.com/blog/155/x-TDL-rootkit–follow-up.html.

[Gla98] Edmund M. Glabus. Metaphors and Modern Threats: Biological, Computer, and Cognitive Viruses. In *7th International Conference and Exhibit Open Source Solutions: Global Intelligence Forum*, March 1998.

[GTA06] Julie Greensmith, Jamie Twycross, and Uwe Aickelin. Dendritic Cells for Anomaly Detection. In *IEEE Congress on Evolutionary Computation*, pages 664–671, July 2006.

[Gur] Lutz Gurtler. Virology of Human Influenza. http://influenzareport.com/ir/virol.htm.

[Hof00] Steven A. Hofmeyr. An Interpretative Introduction to the Immune System. In I. Cohen and L. Segel, editors, *Design Principles for the Immune System and other Distributed Autonomous Systems*. Oxford University Press, April 2000.

[Hua09] Andrew Huang. On Influenza A (H1N1), June 2009. http://www.bunniestudios.com/blog/?p=353.

[Hun93] Lawrence Hunter. *Molecular biology for computer scientists*, pages 1–46. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1993.

[Koi09] Toni Koivunen. Calculating the Size of the Downadup Outbreak, January 2009. http://www.f-secure.com/weblog/archives/00001584.html.

[Kor09] Gert Korthof. Similarities and Dissimilarities of Computer Viruses and Biological Viruses, September 2009. http://home.planet.nl/ gkorthof/korthof78.htm.

[KSA+95] J.O. Kephart, G.O. Sorkin, W.C. Arnold, D.M. Chess, G.J Tesauro, and S.R. White. Biologically Inspired Defenses against Computer Viruses. In *14th International Joint Conference on Artificial Intelligence*, pages 985 – 996, 1995.

[Lov06] Guillaume Lovet. Dirty Money on the Wires: The Business Models of Cybercriminals. In *Proceedings of the Virus Bulletin Conference*, Montréal, Canada, October 2006.

[Man09] Derek Manky. Virut infecting worms, hitching a ride, March 2009.

[Mau11] DDoS Attack on KrebsOnSecurity.com using Russkill Botnet, November 2011. http://www.maurihackers.info/2011/11/ddos-attack-on-krebsonsecuritycom-using.html.

[Mic10] Virus:Win32/Xpaj.gen!A, May 2010. http://www.microsoft.com/security/portal/Threat/Encyclopedia/Entry.aspx?Name=Virus%3AWin32%2FXpaj.gen!A.

[Mik09] Mikko Hypponen. Questions and Answers: Conficker and April 1st, March 2009. http://www.f-secure.com/weblog/archives/00001636.html.

[New10] Researchers Say They Created a Synthetic Cell, May 2010. http://www.nytimes.com/2010/05/21/science/21cell.html.

[NMSF09] Sadia Noreen, Shafaq Murtaza, M. Zubair Shafiq, and Muddassar Farooq. Evolvable Malware, 2009.

[Nob] The Immune System in More Details. http://nobelprize.org/educational_games/medicine/immunity/immune-detail.html.

[NoV10] NoVirusThanks. A New DDoS Bot Named Russkill Is In The Wild, February 2010. http://blog.novirusthanks.org/2010/02/a-new-ddos-bot-named-russkill-is-in-the-wild/.

[Oll10] Gunter Ollmann. Top-10 Botnet Outbreaks in 2009, 2010. http://blog.damballa.com/?p=569.

[PGF93] G. Pantaleo, C. Graziosi, and AS. Fauci. New concepts in the immunopathogenesis of human immunodeficiency virus infection. In *New England Journal of Medicine*, pages 327–335, February 1993. PMID 8093551.

[PM01] Ryan Permeh and Marc Maiffret. Analysis: .ida 'Code Red' Worm, July 2001. http://research.eeye.com/html/advisories/published/AL20010717.html.

[Sec03] WORLD HEALTH ORGANIZATION Secretariat. Smallpox eradication: destruction of Variola virus stocks, 2003.

[SMD+09] SG. Sarafinos, B. Marchand, K. Das, DM Himmel, MA Parniak, SH Hughes, and E. Arnold. Structure and function of HIV-1 Reverse Transcriptase: molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology*, 385:693–713, 2009.

[Sym10] Malicious Code Trends Propagation Mechanisms, 2010. http://www.symantec.com/threatreport/topic.jsp?id= malicious_code_trends&aid= propagation_mechanisms.

[Wika] Discredited AIDS origins theories. http://en.wikipedia.org/wiki/Discredited _AIDS_origins_theories.

[Wikb] Immune System. http://en.wikipedia.org/wiki/Immune_system.

[Wikc] Natural Killer. http://en.wikipedia.org/wiki/Natural_killer.

[Wikd] SARS conspiracy theory. http://en.wikipedia.org/wiki/SARS_conspiracy_theory.

[Wike] Virus. http://en.wikipedia.org/wiki/Virus.

[Wikf] Yellow Fever. http://en.wikipedia.org/wiki/Yellow_fever.

[Wir08] Synthetic Viruses Could Explain Animal-to-Human Jumps, 2008. http://www.wired.com/wiredscience/2008/11/synthetic-virus/.

[Wis] What is the Difference Between a Virus and a Bacteria. http://www.wisegeek.com/what-is-the-difference-between-a-virus-and-a-bacteria.htm.

[Zay09] Vitaly Zaytsev. W32/XPaj: Know Your Polymorphic Enemy, September 2009.