MalSEIRS

An epidemiologic model to predict the behavior of computer viruses

Students of the Applied Mathematics and Computer Sciences (Macc for its acronym in Spanish) program from Universidad del Rosario developed a research project on the application of epidemiologic models to understand malicious software (MalSEIRS). With this, they want to offer better solutions to face the cyberattacks companies are exposed on a daily basis.

By Camilo Calderón Acero

Photos Milagro Castro, 123RF .
DOI https://doi.org/10.12804/dvcn_10336.36921_num6

In November 2021, the website of the Departamento Nacional de Estadísticas (National Department of Statistics, Dane) had to be taken down for a couple of hours, which removed access to hundreds of users who use this platform daily. The reason was a cyberattack against the entity. Cybercriminals deleted the databases containing sensitive and confidential information, affecting around 420 servers within the network.

We can say that the network was "infected" with a virus that made all its components "sick," just as when COVID-19 spread throughout the world and affected human beings. While COVID-19 is airborne and spreads in close contact, the Dane systems were infected because the equipment was connected to the Internet.

As in the example, everyday equipment, programs, or apps, devices, or platforms (IT assets) connected to the network are exposed to multiple risks, and measures must be taken to prevent them. Something similar happened during the pandemic, when quarantines and the use of facemasks were ordered, frequent hand wash was suggested, vaccines were being developed in record time, and later, intense vaccination sessions were implemented to counterbalance the spread of numerous SARS-CoV-2 subvariants.

The similarities are no coincidence. Students of the Applied Mathematics and Computer Sciences (Macc for its acronym in Spanish) program from Universidad del Rosario (URosario), Isabella Martínez Martínez, and Andrés Felipe Florián, together with researchers from Universidad de Murcia in Spain and coordinated by Daniel Díaz López, Principal Professor at the School of Engineering, Science and Technology of URosario, discovered that they could apply the parameters of a detailed

↑ A key point in the study performed by the group of students was to determine how similar the model proposed is to other predictive models. One of the most important findings here was the more rapid stabilization of the MalSEIRS model.

epidemiologic model to the cybersecurity field. Thus, they could pinpoint the behavior of malicious software (malware) and proposed formulas (algorithms) to avoid them. The results were published in Complexity in December 2021.

They were mainly based on the SEIRS model—susceptible, exposed, infected, recovered, and susceptible again—a complex, yet also very precise method commonly used in the epidemiology field that considers a great spectrum of scenarios in which a viral attack may happen and an infectious disease may progress. In the recent pandemic, its application was of great use to establish infection peaks—that is, the periods with more positive cases confirmed.

Malware is one of the most lethal threats in cyberspace. It is known for corrupting and damaging IT systems silently. It can steal, encrypt, and delete data as well as spy on the user's activity without nobody noticing. "It is like a disease, a condition that poses IT assets under permanent threat. As happens with breathing and talking to other people without a facemask, when we run the risk of catching a respiratory disease, computers communicate with each other via networks, and this is how they can spread malware," explains Isabella.

Epidemiology applied to cybersecurity

The researchers built their mathematical model based on epidemiology's own methodologic resource. To achieve this, it was essential to analyze the behavior of cyberthreats from preexisting mathematic assumptions, a field of study that has been developing globally for around 20 years.

"Our research is framed in the context of infectious diseases and, in this field, its most basic epidemiologic model is SIR (susceptible - infected - recovered), which takes data from people that meet certain conditions (that is, the number of cases that go from one stage to the other, for example, from being infected to being recovered)," explains Martínez.

"It is a very simple model that handles statistical rates. This means that the rate at which individuals (or computers, in our case) move from one denomination to the other (from susceptible to exposed, for example) does not change in time. For instance, if the rate of susceptible and infected people is 0.5, this means that, in a given time, half of the susceptible people will become infected."

Each parameter in this model allows us to make an inference over the course of the disease. For example, if there is a high infection rate and a high recovery rate, we can say the virus is infectious, but the disease it causes is not severe. There is where mathematics is useful to understand the scope of a pathology or, in this case, a cyberthreat.

However, SIR is a simple model since it does not consider new subgroups of "subjects" (IT equipment) that may appear when handling large quantities of infected equipment (over 100 units) and when infection has been present for over 1 day. For example, the Exposed-to-the-Virus subgroup (E) represents the nodes before being infected and may spread the disease, or the Susceptible subgroup (S), where nodes lose previously acquired immunity.

The SEIRS model does consider these new subgroups (E and S), so it is a more complex model compared to SIR. But, at the same time, it is more precise since it allows for the inclusion of new scenarios in which a cyberattack may happen.

When applying the analogy of epidemiology to the universe of cybersecurity and malware, we obtain the acronym MalSEIRS, a model that considers a larger number of attack parameters (i.e., changes in the incubation rate, deaths, and population size) and that these parameters may vary in time. In the MalSEIRS model, considering progress in time is vital since it better adjusts to what really happens during a cyberattack.

"Rates vary in time because nobody is aware of the infection when it's starting, and the virus spreads massively. When we are aware of the infection, and we start protecting ourselves with the proper supplies or mechanisms, the rate starts dropping. The same happens with computers. When a malware is known to be circulating through systems, a protection mechanism is placed on the susceptible nodes, which makes infection rate to drop," states Díaz López.

I
R
S
Infected
Recovered
Susceptible

In this respect, Florián adds the following: "What we do is take rates from the model that depend on time—infection, recovery, propagation, and immunity loss—and extrapolate them to the malware infection events. What happens with the infection rate is that the more pieces of equipment are infected, the less "healthy" pieces available there will be to be infected (and, as a secondary effect, more network congestion may happen). Therefore, the infection rate will vary: it will rise in the early infection stages, and drop in the later stages."

The great restriction of the previous models is that they do not consider that the variation rates among categories may vary in time. We need to consider that as a cyberattack progresses, the human and technological team in charge of incident response will take containment measures to diminish the infection rate or make it drop to zero. The same applies for other rates in the model that may also vary in time for external reasons.

For Professor Carlos Arturo Castillo Medina, Director of the Graduate Diploma in Telematic Network Security of Universidad El Bosque, it is increasingly necessary to have this kind of tools that allow organizations to make timely decisions since cyberattacks are usually more common and lethal than people think.

"The great strength of this project is the malware behavioral study, which allows for the development of profiles based on suspicious behaviors and behavior guidelines of the infectious cyberagent. This enables its detection before we have seen its attack or course of action," he added.

From theory to practice

A key point of the study performed by the Macc students was to determine the similarities between the model proposed and other predictive models. One of the most important findings here was the more rapid stabilization of the MalSEIRS model. This means that their figures could assimilate those in reality more rapidly than in other models.

→

Daniel Díaz López, professor at the School of Engineering, Science and Technology of URosario.

To reach such conclusion, they first reviewed the ranges of each parameter in the model individually to obtain data of their application at a large scale. Then, the results were compared with similar models. To justify the variability of rates in the model, they used data available from other attacks, such as those committed by the *Wannacry* worm, in 2017; the *Slammer* worm in 2003; and the *Emotet* Trojan, discovered in 2014.

The third phase of the project assessed defense mechanisms that may be implemented to stabilize the model and reach the required rates to contain an attack in a robust way. “We thoroughly reviewed parameters such as vaccination or initial susceptibility rates for the devices entering the network,” says Martínez.

“We analyzed how high the rates must be to contain a malware in a network. This value in itself is a defense mechanism, because it lets us know how many pieces of equipment must be immunized or how much money we must invest in antivirus licenses to get rid of the malware.”

Simulations considered several scenarios, such as computers losing the “antivirus signature” (the file that tells the antivirus software to find risks and repair the threatened systems) after a certain amount of time passed, letting new pieces of equipment into the network with a probability of being susceptible or a susceptible piece of equipment being recovered quickly.

Data to make *cyberdecisions*

Cybersecurity is not a minor topic in Colombia. Besides the Dane attack, in 2021, over 20,500 cybercrime reports were submitted before the Prosecutors Office. In fact, it is estimated that, during said year, there were 87 threat attempts per minute according to data from the report titled *The Scenario of Threats in Latin America 2021.*

Thus, for researchers, it was important that the *MalSEIRS* model could provide valuable information and recommendations on defense and attack mechanisms to organizations. Therefore, the findings from the research were compiled in a playbook that can be offered as a handbook for companies to know what to do in case they are victims of a cyberattack (defense strategies) or what measures can be implemented to preventively attack possible threats of this kind (attack strategies).

Currently, not all organizations have a structure to react to these incidents and understand the nature of an attack, which can characterize it properly and apply the ideal defense mechanism. What usually happens is that reactive measures are taken from the moment the incident happens.

As in the Dane event, Internet disconnection (which would be the obvious choice) is not always the best alternative. “Disconnecting” implies economic or reputation loss for the company, which may even be bigger than the loss caused by the malware. For this reason, we must think on intermediate defensive strategies that allow for operating in a safer way and, at the same time, keep potential threats under control.

The researchers’ proposal for this scenario is to apply different measures, such as allowing the network to only connect new equipment that has been previously immunized. “They must be immunized before acting so they are not exposed to infection. The second thing I would personally do is isolate all the devices infected to prevent them from spreading the malware further. It is also essential to implement immediate ‘vaccination’ measures in equipment within the network, since once connected they are susceptible to the risk, and also treatment measures for the infected equipment,” comments professor Díaz López.

For the project members, it is valuable that mathematics and epidemiology can offer solutions for one of the greatest problems that companies face today.

Types of malware cyberattacks

Malware: malicious software that can invade operating systems and cause all kinds of damages (steal information, cause damages to the device, obtain an economic benefit, take control of the device, etc.).

Virus

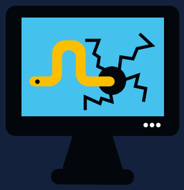
Designed to copy themselves and propagate to as many devices as possible. They use transport means such as external memories or electronic mail.

Spyware

Their objective is spying and stealing information. They are also capable of downloading other malware and installing them in the device.

Worm

Its objective is to multiply, creating copies of itself and spreading across the network. Unlike viruses, they require no action from the user.

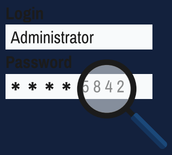


Trojans

Once they enter the system, their objective is to create an access point for harmful software to enter. They are usually disguised as legitimate software.

Keyloggers

They record and follow up on every key pressed in a device without our consent. They may be based in a software or hardware, such as a USB device.



Ransomware

They have a bigger impact, particularly economic, in users. Their objective is data kidnapping to demand a ransom in exchange of not disclosing or destroying said data.

Adware

Designed to show unwanted ads in a massive way. They cause little harm but are quite bothersome for users.

Backdoors

They will allow the cybercriminal to take control of the device remotely. They are commonly used to infect several devices and create a zombie network, or Botnet.

Definitions from:
Guideline of cyberattacks from the Instituto Nacional de Ciberseguridad (National Institute of Cybersecurity) (INCIBE) de España https://www.incibe.es

MaISEIRS is one of the most effective tools right now. There are challenges within their work since this model can only be applied on existing malwares and new threats whose behavior may not be predicted by the model appear on a daily basis. It is also very common that organizations do not monitor cyber risks properly, so there usually is no reliable, real-time data that allow making timely decisions.

In this sense, engineer Castillo warns that although these prediction tools are an important contribution to the battle against cybercriminals, they also have limitations, such as showing 'false positive' results. "Some data can be interpreted as malware when it is probably just a software searching for updates. This requires a second revision from another perspective. My having a headache, bone pain and a fever does not necessarily mean that I caught a cold. It may be laryngitis, tonsilitis or even COVID-19," he comments.

→ Isabella Martínez, student of the Applied Mathematics And Computing Sciences (Macc) program at Universidad del Rosario.

→ Andrés Felipe Florián, student of the Applied Mathematics And Computing Sciences (Macc) program at Universidad del Rosario.

Both Florián and Martínez, the two main authors, agree that the study is more precise in the field of cybersecurity, combining theory and practice. Both want the foundations to go beyond mathematics – to use concepts from cybersecurity and forensic analysis to be applied to real-life situations and, thus, offer new and practical solutions.

Both Florián and Martínez, the two main authors, agree that the study is more precise in the field of cybersecurity, combining theory and practice. Both want the foundations to go beyond mathematics to use concepts from cybersecurity and forensic analysis to be applied to real-life situations and, thus, offer new and practical solutions. ◘