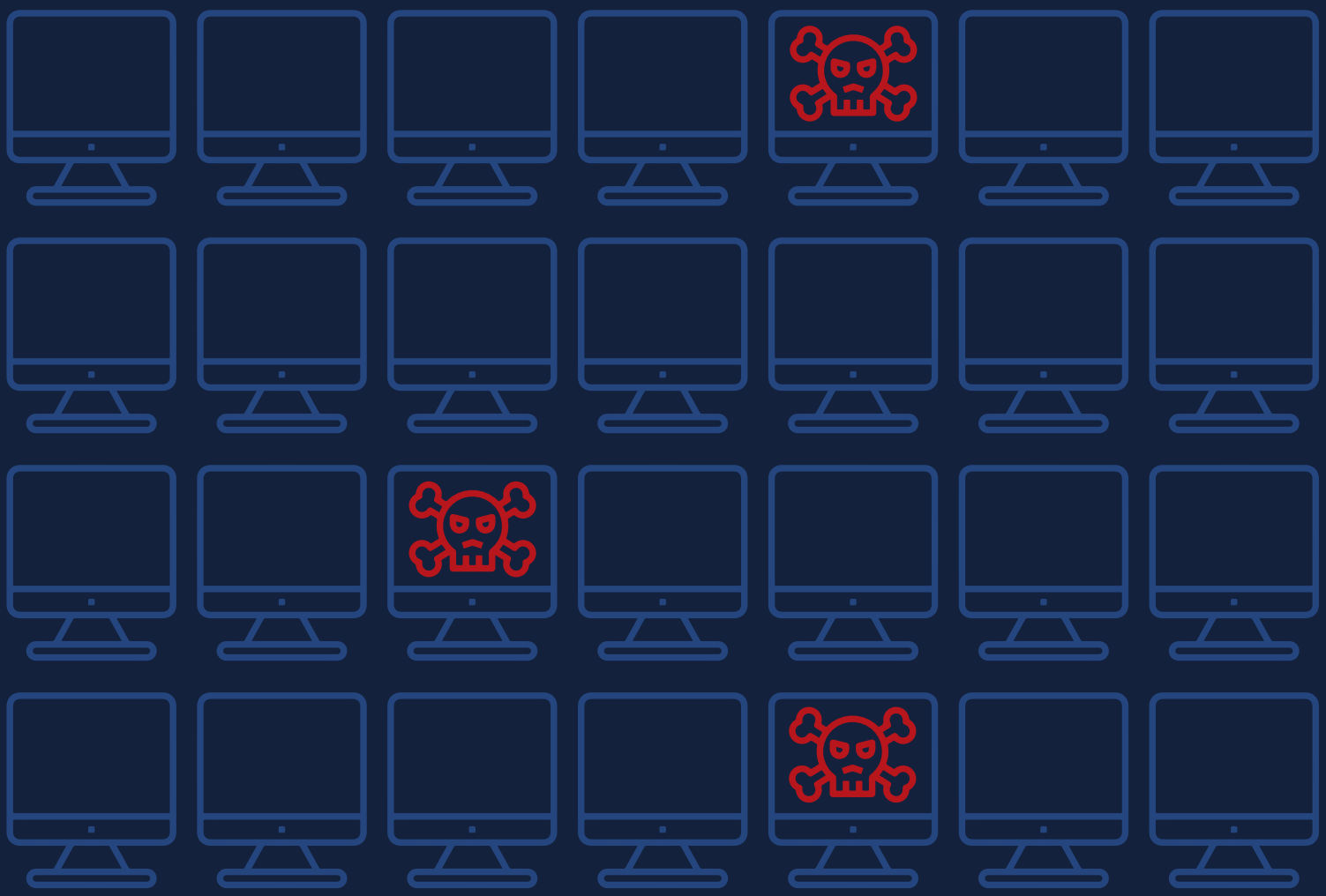


MalSEIRS

Un modelo epidemiológico para predecir el comportamiento de virus informáticos

Estudiantes del programa de Matemáticas Aplicadas y Ciencias de la Computación (Macc), de la Universidad del Rosario, desarrollaron un proyecto de investigación sobre la aplicación de modelos epidemiológicos para entender el software malicioso (MalSEIRS). Con él pretenden ofrecer mejores soluciones para enfrentar los ataques informáticos a los que se exponen a diario las empresas.

Por Camilo Calderón Acero
Fotos Milagro Castro, 123RF .
DOI https://doi.org/10.12804/dvcn_10336.36921_num6

En noviembre de 2021 el portal web del Departamento Nacional de Estadística (Dane) tuvo que ser desactivado un par de horas, lo que dejó sin acceso a cientos de usuarios que requieren a diario esta plataforma. La razón del bloqueo obedeció a que la entidad sufrió un ataque informático. Ciberdelincuentes borraron sus bases de datos con información de carácter reservado, sensible y confidencial, con lo cual se afectaron alrededor de 420 servidores de su red.

Se podría decir que esta se 'infectó' con un virus y 'enfermó' a todos sus integrantes, así como cuando la COVID-19 se extendió por el mundo y afectó a los seres humanos. En este caso la transmisión se da por el contacto cercano vía aérea, pero en el de los sistemas del Dane la infección fue posible porque los equipos estaban conectados a internet.

Tal como en el ejemplo, a diario los equipos, los programas o aplicaciones, los dispositivos o plataformas (activos informáticos) que están conectados a la red se ven expuestos a múltiples riesgos y se debe tomar medidas para que esto no pase. Algo muy similar ocurrió durante la pandemia, cuando se ordenaron confinamientos, uso obligatorio de tapabocas, lavado frecuente de manos, se desarrollaron vacunas en tiempo récord, y se adelantaron intensas jornadas de vacunación para contrarrestar la diseminación de las numerosas variantes del SARS-CoV-2.

Esta similitud no es solo una coincidencia. Los estudiantes del programa de Matemáticas Aplicadas y Ciencias de la Computación (Macc) de la Universidad del Rosario (URosario), Isabella Martínez Martínez y Andrés Felipe Florián, junto con investigadores de la universidad de Murcia en España y bajo la coordinación de Daniel Díaz López, profesor principal de la Escuela de Ingeniería, Ciencia y Tecnología de la URosario, descubrieron que se podían aplicar los parámetros de un modelo epidemiológico detallado al campo de la ciberseguridad. Así pudieron precisar el comportamiento de un *software* malicioso (*malware*) y plantearon fórmulas (algoritmos) para su contención. Los resultados fueron publicados en la revista *Complexity* en diciembre de 2021.

Se basaron principalmente en el modelo SEIRS (susceptible, expuesto, infectado, recuperado y –de nuevo– susceptible), un complejo pero al mismo tiempo muy preciso método utilizado comúnmente en el universo de la epidemiología, que considera un gran espectro de escenarios en los que se puede dar un ataque viral y por los que una enfermedad infecciosa puede avanzar. En la reciente pandemia su aplicación fue de gran utilidad para establecer los picos del contagio, es decir, los periodos con más casos positivos confirmados.

El *malware* es una de las amenazas más letales que existen en el ciberespacio. Se caracteriza por corromper y dañar los sistemas informáticos de forma silenciosa. Puede robar, cifrar, borrar datos y espiar la actividad del usuario sin que nadie lo advierta. "Es como una enfermedad, una condición que pone en permanente riesgo los activos informativos. Así como al respirar y al hablar con otras personas sin la protección de un tapabocas corremos el riesgo de contraer alguna infección respiratoria, los computadores se comunican entre ellos por redes y de esa manera pueden transmitir el *malware*", explica Isabella.

Epidemiología aplicada a la ciberseguridad

Los investigadores construyeron su modelo matemático basados en dicho recurso metodológico propio de la epidemiología. Para lograrlo fue indispensable analizar el comportamiento de las amenazas informáticas a partir de postulados matemáticos ya existentes, un campo de estudio que lleva alrededor de 20 años de avance en el mundo.

"Nuestra investigación se remonta al contexto de las enfermedades infecciosas y en ese campo el modelo epidemiológico más básico que hay es el SIR (susceptibles-infectados-recuperados), el cual toma datos de la tasa de las personas que cumplen dichas condiciones, es decir, la cantidad de casos que pasan de un estadio a otro, por ejemplo, de estar infectados a estar recuperados", explica Martínez.

"Es un modelo muy sencillo que maneja tasas estáticas. Esto significa que la tasa a la que se mueven los individuos (o los ordenadores para nuestro caso) entre una denominación y otra -de susceptible a expuesto por ejemplo- no cambia en el tiempo. Por ejemplo, si la tasa de personas susceptibles e infectadas es de 0,5, esto quiere decir que en un tiempo determinado, la mitad de los susceptibles se van a infectar".

Cada uno de los parámetros de este modelo permite establecer una inferencia sobre el curso de la enfermedad. Por ejemplo, si la tasa de infección es alta, pero también es elevada la de recuperación, se puede decir que el virus es infeccioso, pero la enfermedad que produce no es grave. Es allí donde las matemáticas ofrecen su utilidad para comprender el alcance de una patología o, como en este caso, el de una amenaza informática.

Un punto trascendental del estudio realizado por el grupo de estudiantes fue determinar qué tanto se asemeja el modelo propuesto a otros modelos predictivos. Uno de los hallazgos más destacables en este punto fue la estabilización más acelerada del modelo MalSEIRS.

Sin embargo, el SIR es un modelo simple, pues no considera nuevos subgrupos de 'sujetos' (equipos informáticos) que pueden aparecer cuando se manejan cantidades considerables de equipos infectados (más de 100) y cuando la infección ha estado presente por un tiempo superior a 1 día. Por ejemplo, el subgrupo de Expuestos al virus (E) que representa a los nodos previos a ser infectados y que pueden transmitir la enfermedad o el subgrupo de Susceptibles (S), que representa nodos que pierden la inmunidad previamente conseguida.

El modelo SEIRS si considera estos nuevos subgrupos (E y S), siendo un modelo más complejo en comparación con SIR, pero al mismo tiempo más preciso, pues permite incluir nuevos escenarios en los que se puede dar un ataque.

Al aplicar la analogía del área de la epidemiología al universo de la ciberseguridad y el *malware,* obtenemos el acrónimo MalSEIRS, un modelo que considera una mayor cantidad de parámetros de los ataques (p. ej. cambios en la tasa de incubación, de muertes y del tamaño de la población) y que estos puedan variar en el tiempo. En el modelo MalSEIRS la consideración del progreso en el tiempo es vital, pues se ajusta más a lo que ocurre realmente en un ataque informático.

"Las tasas varían en el tiempo porque al inicio de la infección nadie es consiente de ella y el virus se propaga de forma masiva. Cuando ya somos conscientes de la enfermedad y nos comenzamos a proteger con los insumos o los mecanismos adecuados, la tasa comienza a bajar. Con los computadores pasa lo mismo. Cuando ya se conoce que un *malware* está circulando en los sistemas, se empieza a generar un mecanismo de protección sobre los nodos susceptibles que hacen que la tasa de infección baje", indica Díaz López.

Al respecto, Florián complementa: "Lo que hacemos es tomar las tasas del modelo dependientes del tiempo –de infección, de recuperación, de propagación y de pérdida de inmunidad – y extrapolarlas a los eventos de infección del *malware*. Con la tasa de infección sucede que entre más equipos se empiecen a infectar, menos equipos 'sanos' disponibles habrá para ser infectados (además de que, como efecto colateral, se puede presentar mayor congestión en la red). Entonces, la tasa de infección variará: se elevará en los estados tempranos de la infección y se reducirá en los más tardíos".

La gran restricción de los modelos anteriores es que no consideran que las tasas de variación entre categorías puedan variar en el tiempo. Hay que pensar que a medida que un ataque informático avanza el equipo humano y tecnológico de respuesta a incidentes tomará medidas de contención destinadas a que disminuya –o llegue a cero– esa tasa de infección. Lo mismo aplica para las otras tasas del modelo que también pueden variar en el tiempo por razones externas.

Para el profesor Carlos Arturo Castillo Medina, director de la especialización en Seguridad de Redes Telemáticas de la Universidad El Bosque, cada vez es más necesario contar con herramientas como estas que permitan a las organizaciones tomar decisiones oportunas, ya que los ataques informáticos suelen ser más comunes y letales de lo que se cree

"La gran fortaleza que tiene este proyecto es el estudio comportamental del *malware*, el cual permite el desarrollo de perfiles basados en comportamientos sospechosos y pautas de conducta del agente informático infeccioso. Así es posible detectarlo antes de haber observado su ataque o forma de actuar", añade.

Daniel Díaz López, profesor de la Escuela de Ingeniería, Ciencia y Tecnología de la URosario.

De la teoría a la práctica

Un punto trascendental del estudio realizado por el grupo de estudiantes del Macc fue determinar qué tanto se asemeja el modelo propuesto a otros modelos predictivos. Uno de los hallazgos más destacables en este punto fue la estabilización más acelerada del modelo MalSEIRS. Esto quiere decir que sus cifras lograron asimilarse a las de la realidad más pronto que otros modelos.

Para llegar a dicha conclusión primero se revisaron los rangos de cada uno de los parámetros del modelo, de forma separada, con el fin de obtener datos de su aplicación a gran escala. Luego se hicieron comparaciones de resultados con modelos similares. Para justificar la variabilidad de las tasas del modelo se usaron datos disponibles de otros ataques como los perpetrados por los gusanos *Wannacry*, en 2017; *Slammer* en 2003; y el troyano *Emotet*, descubierto en 2014.

En la tercera fase del proyecto se evaluaron mecanismos de defensa que se podrían implementar para estabilizar el modelo y así lograr las tasas requeridas para contener un ataque de manera contundente. "Revisamos minuciosamente parámetros como la tasa de vacunación o la de susceptibilidad inicial de los dispositivos que ingresan a la red", indica Martínez.

"Analizamos qué tan altas debían estar para contener un *malware* en una red. Este valor, en sí, es un mecanismo de defensa porque nos permite conocer qué tantos equipos se deben inmunizar o qué tanto debo invertir en licencias de antivirus para deshacerme del *malware*".

Las simulaciones consideraron varios escenarios, como por ejemplo que los computadores pierdan la 'firma antivirus' (el archivo que informa al *software* antivirus para encontrar los riesgos y reparar los sistemas amenazados) luego de que transcurra cierta cantidad de tiempo, que ingresen nuevos equipos a la red con una probabilidad de que sean susceptibles o que un equipo que sea susceptible pueda pasar a ser recuperado rápidamente.

Datos para tomar *ciberdecisiones*

El tema de la ciberseguridad no es menor en Colombia. Además del ataque al Dane, en 2021 se presentaron más de 20.500 reportes criminales ante la Fiscalía por ciberdelitos. De hecho, se estima que en ese mismo año ocurrieron 87 intentos de amenaza por minuto, según datos del informe *El panorama de amenazas en América Latina 2021.*

De allí que para los investigadores era importante que a partir del modelo *MalSEIRS* se pudiera brindar información de valor y recomendaciones sobre mecanismos de defensa y de ataque a las organizaciones. Por ello el conocimiento surgido de la investigación se compiló en un *playbook* (libro de jugadas) que se puede ofrecer a manera de manual a las empresas para que sepan qué se debe hacer en caso de ser víctimas de un ciberataque (estrategias de defensa) o qué se puede implementar para atacar de manera preventiva posibles amenazas de este tipo (estrategia de ataque).

En la actualidad, no todas las organizaciones cuentan con una estructura de respuesta a dichos incidentes que pueda entender la naturaleza de un ataque, caracterizarlo debidamente y aplicar el mecanismo de defensa idóneo. Lo que suele suceder es que se toma una medida reactiva a partir del momento en que ocurre el incidente.

Tal como en el evento del Dane, la desconexión de internet –que sería la elección obvia– no siempre es la mejor alternativa. 'Desconectar' implica una pérdida económica o reputacional para la empresa, que incluso puede ser más cuantiosa que la causada por el *malware*. Por esa razón deben plantearse estrategias defensivas intermedias que permitan funcionar de manera segura, al mismo tiempo que mantener a raya las potenciales amenazas.

La propuesta de los investigadores para este escenario es aplicar medidas diferentes, tales como permitir que se conecten a la red solo nuevos equipos previamente inmunizados. "Se deben inmunizar antes de entrar en acción para que no estén expuestos a una infección. Lo otro que personalmente haría es aislar por completo los dispositivos que estén infectados para evitar que sigan propagando el *malware.* Así mismo, es indispensable implementar medidas de 'vacunación' inmediata en los equipos que están dentro de la red, pues una vez conectados ya son susceptibles al riesgo, y también medidas de tratamiento para los que ya están enfermos", asegura el profesor Díaz López.

Para los integrantes del proyecto resulta valioso que las matemáticas y la epidemiología puedan ofrecer soluciones a una de las grandes problemáticas que enfrentan las empresas hoy

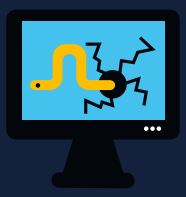
Tipos de ciberataques por *malware*

Malware: *software* malicioso capaz de invadir sistemas operativos y causar todo tipo de daños (robar información, causar daños en el equipo, obtener un beneficio económico, tomar control del equipo ,etc).

Virus
diseñados para copiarse a sí mismos y propagarse a tantos dispositivos como les sea posible. Utilizan medios de transporte como memorias externas o correo electrónico

Spyware
Su finalidad es espiar y robar información. También es capaz de descargar otros *malware* e instalarlos en el equipo.

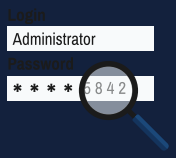
Gusano
Tiene como objetivo multiplicarse creando copias de sí mismo, distribuyéndose por toda la red. A diferencia de los virus, estos no requieren ninguna acción por parte del usuario para ejecutarse.



Troyanos
Una vez que ingresa tiene el objetivo de crear un acceso para que puedan ingresar otros *software* dañinos. Suelen camuflarse como un *software* legítimo.

Keyloggers

Realizan un seguimiento y registran cada tecla que se pulsa en un equipo sin nuestro consentimiento. Pueden estar basados en un *software* o en un *hardware*, como por ejemplo un dispositivo USB.



Ransomware

De mayor impacto, especialmente económico en los usuarios, ya que su objetivo es el secuestro de datos para exigir rescate a cambio de no hacerlos públicos o no destruirlos.

Adware

Diseñados para mostrarnos anuncios no deseados de forma masiva. Causan poco daño pero son muy molestos para los usuarios

Backdoors

Permitirá al ciberdelincuente tomar el control del equipo de forma remota. Suelen utilizarse para infectar a varios dispositivos y formar una red zombi o Botnet.

Definiciones basadas en:
Guia de ciberataques del Instituto Nacional de Ciberseguridad (INCIBE) de España https://www.incibe.es

en día. MaISEIRS es una de las más efectivas en el momento. Quedan retos en su trabajo, en el sentido de que este modelo solo se puede aplicar en *malwares* que ya existen y teniendo en cuenta que cada día surgen nuevas amenazas cuyo comportamiento no podría ser predicho por el modelo. También es muy común que las organizaciones no hagan una buena monitorización de los ciberriesgos, por lo que no suele haber datos confiables y en tiempo real que permitan tomar decisiones oportunas.

En ese sentido, el ingeniero Castillo advierte que estas herramientas de predicción, aunque generan una importante contribución a la lucha contra los ciberdelincuentes, también tienen limitaciones como, por ejemplo, arrojar 'falsos positivos'. "Algunos datos se pueden interpretar como *malware* cuando posiblemente sean solo un *software* en búsqueda de actualizaciones. Eso plantea una segunda revisión desde otra perspectiva. No porque me duelan la cabeza y los huesos, y tenga fiebre quiera decir, necesariamente, que sufro un episodio de gripa. Puede ser laringitis, amigdalitis o incluso COVID-19", reflexiona.

Tanto Florián como Martínez, los dos autores principales, concuerdan que el estudio brinda una mayor precisión en el campo de la ciberseguridad, combinando lo teórico y lo práctico. Ambos buscan que los fundamentos se queden no solo en las matemáticas, sino que se usen con conceptos de ciberseguridad y análisis forense para que puedan aplicarse a casos de la vida real, y así ofrecer soluciones prácticas y novedosas. ■

→ Isabella Martínez, estudiante del programa de Matemáticas Aplicadas y Ciencias de la Computación (Macc) de la Universidad del Rosario.

→ Andrés Felipe Florián, estudiante del programa de Matemáticas Aplicadas y Ciencias de la Computación (Macc) de la Universidad del Rosario.

Tanto Florián como Martínez, los dos autores principales, concuerdan que el estudio brinda una mayor precisión en el campo de la ciberseguridad, combinando lo teórico y lo práctico. Ambos buscan que los fundamentos no solo se queden en las matemáticas, sino que se usen conceptos de ciberseguridad y análisis forense que puedan aplicarse a casos de la vida real y así ofrecer soluciones prácticas y novedosas al sector.