Antivirus Technology Offers New Cures

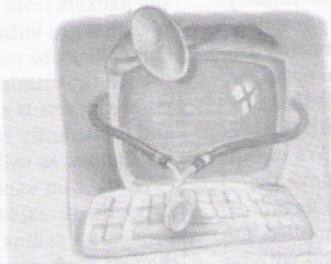
Lee Garber and Richard Raucci

Work on computer-virus theory began 50 years ago, when John von Neumann described self-replicating systems.

Since then, viruses have become a serious security threat to casual home computer users and large corporate networks alike. The effort to combat this threat has spawned an entire industry.

Over the years, the antivirus industry has had to keep pace as virus writers have become more sophisticated. Antivirus products now not only detect and eliminate viruses, they can even delete or repair infected files, and remove infected sectors from system memory and disk drives.

Researchers have taken many approaches, and some of the newest and most promising antivirus technology is modeled on the way the human body fights viruses.



Some of the newest and most promising antivirus technology is modeled on the way the human body fights viruses.

BIOLOGICAL MODELS

Researchers have been looking into biological models for computer antivirus systems for several years. Some of this research is based on the similarities between human and computer viruses. Both types of viruses latch onto a host, use its resources to reproduce, and cause a range of symptoms.

Stephanie Forrest, associate professor at the University of New Mexico's Computer Science Department noted that in her antivirus research, "We were making models of the immune system and thinking about the immune system from an information processing point of view."

Editor: Lee Garber, *Computer*, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1314; l.garber@computer.org

Immune System for Cyberspace

One of the most elaborate biologically based systems is IBM's Immune System for Cyberspace (http://www.av.ibm.com/current/FrontPage/), shown in the figure on the next page.

The Immune System will be linked to customers' computer networks via the Internet. It detects viruses using two heuristics, said Jeffrey O. Kephart, manager of IBM's agents and emergent phenomena group

One heuristic uses statistical techniques to create byte-signature fingerprints of uninfected executable files. The system then uses pattern-matching algorithms to compare these fingerprints to subsequent fingerprints of the files, Kephart explained. Changes indicate a possible viral infection. The Immune System then analyzes the changes to determine whether they were actually caused by a virus.

According to Kephart, the Immune System can also use a neural-network technique that quickly identifies short byte sequences that represent instructions for carrying out virus-related tasks. Programs with several of these sequences probably have a viral infection, he said. This technique can detect previously unknown viruses that include these viral instruction patterns, which is an advantage over systems that can recognize only known viruses, he added.

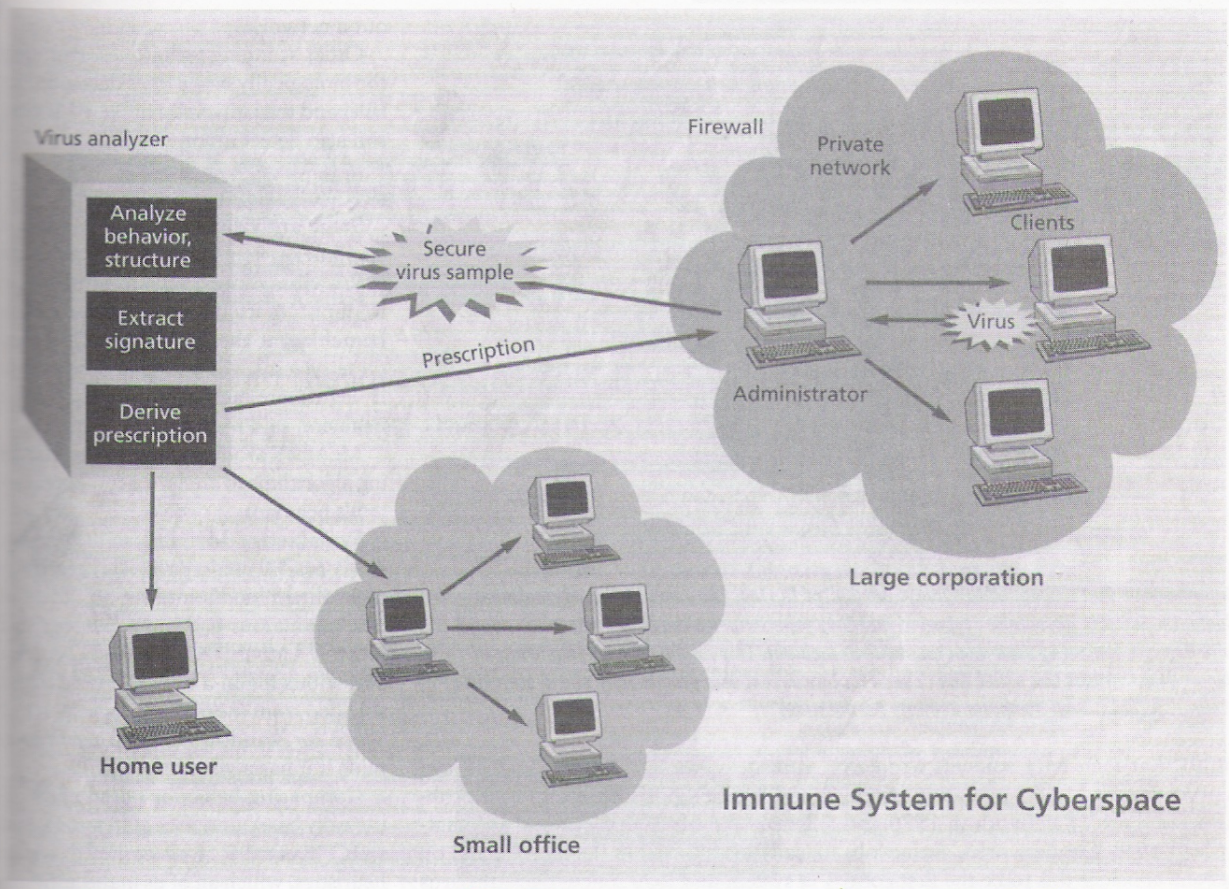
Once an infection has been identified, a copy of the infected program is sent to a central IBM computer. The computer's software provokes a virus into action on a decoy program so that it can be analyzed. The system then compares infected and uninfected decoy programs, using pattern-matching algorithms, to determine a virus' structure and the way it causes infections, Kephart said.

The central computer then develops, tests, and distributes a cure to the infected machine and all computers on the same network. The cure, which would immunize computers against future infection by the same virus, could also be distributed to other organizations that use the Immune System, Kephart said.

In tests, he said, the Immune System completed this process automatically in about three minutes and, in commercial settings, could do so "certainly in less than 10 minutes."

In many ways, the Immune System functions like the human body, which also must recognize that an infection

Reprinted from *IEEE Computer*, Vol. 31, No. 2, pp. 12–14, June 1997.



IBM's Immune System for Cyberspace uses two types of heuristics to detect viruses. Once a virus is detected in an organization's machine, it is sent to the Immune System. The Immune System provokes the virus into action on an isolated decoy program to study how it works. The system then develops, tests, and distributes a cure to the infected computer, to other machines on the same network, and potentially to other Immune System customers.

agent is present, identify the agent, and then develop and implement a cure. And like the human body, IBM's technology also remembers and stores antibody information, which reduces response time to future infections.

IBM plans to begin a pilot program for selected clients in the near future and to sell Immune System services commercially by the middle of this year, Kephart said.

T-cell model

The University of New Mexico's Forrest has modeled a biologically based approach to antivirus technology on another aspect of the human immune system: T cells.

T cells have many different receptors on their surface, each of which can bind to a different foreign material the cell has not contacted before. This helps the body identify and subsequently fight harmful foreign materials in the body.

Forrest said the algorithm she developed generates random byte patterns, which function like the many receptors on the surface of T cells. The algorithm compares each pattern against existing code in a file, and if there is no match, it stores the pattern. If the pattern shows up later, that means the program has changed. This step is the equivalent of the way the human body uses T cells to identify foreign materials. Forrest said her algorithm uses a partial-matching rule with a threshold that makes sure only changes of a certain extent and nature are analyzed.

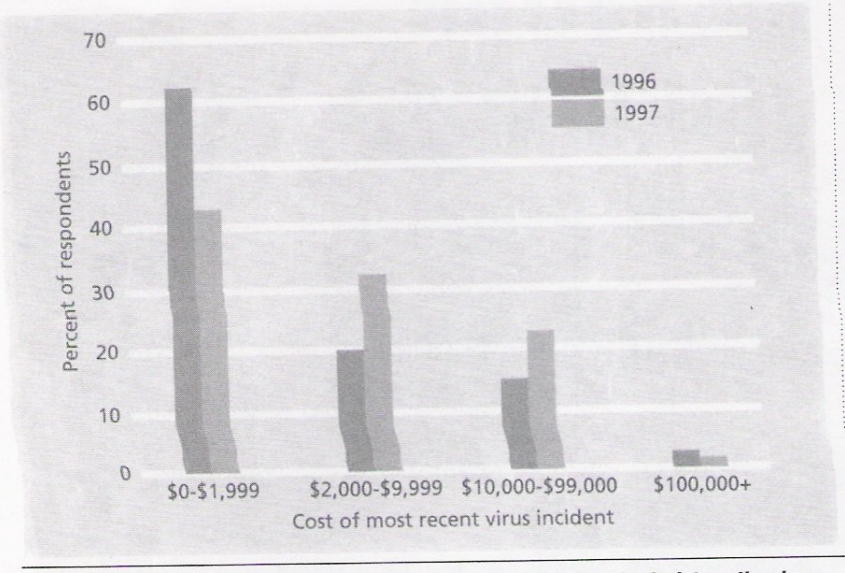
Her algorithm is not designed to specifically identify viruses but to determine changes in files, programs, and patterns of activity that could indicate the presence of a virus or some other problem. She said the algorithm will recognize changes even if caused by a virus the system hasn't seen before.

Because her algorithm is based on the human immune system, which is distributed and parallel, she said, her algorithm makes the most sense in computer settings where you also want distributed protection, such as networked systems or multiple computers that are running the same software. She emphasized that the algorithm is not a full antivirus system but could be used in one.

HISTORY AND PREVALENCE

The development of biological antivirus technologies is the latest step in the history of computer viruses, which began 50 years ago with von Neumann's theories on self-replicating systems.

The first attempts to write programs based on von Neumann's work occurred in the 1960s. A self-replicating program called "Cookie Monster" (ftp://ftp.stratus.com/pub/vos/multics/tvv/cookie.



Organizations that responded to surveys conducted the last two years by the International Computer Security Association reported that they experienced a range of costs while coping with damage caused by their most recent virus incident. This reflects the varying levels of damage that viruses can cause due to such factors as lost user productivity and IS staff time. (Source: International Computer Security Association)

html) that some MIT students wrote as a prank for the Multics operating system affected the MIT network in 1970 and spread to other networks. Some observers consider this to be the first computer virus.

Viruses began affecting desktop computers in the 1980s, starting with a harmless virus that infected Apple II systems in 1981. Viruses affected PCs and the Arpanet shortly thereafter. Since the late 1980s, the number of computer viruses has increased rapidly. While some viruses can be relatively benign, some can destroy files and even hard drives, with devastating effects on an organization. The figure above demonstrates the range of damage that viruses can cause.

Last December, the WildList information service (http://www.virusbtn.com/WildLists/199712.html), which works with 46 virus researchers reported that it found 256 different viruses in action currently throughout the world.

The Norton AntiVirus products by Symantec track about 13,000 different virus signatures. Dr. Solomon's AntiVirus Toolkit tracks more than 15,000 virus fingerprints.

VIRUSES AND THE INTERNET

Viruses have become a particular risk with the advent of Internet technology, which makes it much easier to pass and spread viruses. Such technologies as Java and ActiveX pose important new problems. An ActiveX applet could automatically install itself in the background on a system and load a new file. Antivirus researchers fear a hostile applet could introduce a virus that way.

New viral threats will probably be along these lines, said Shannon Talbott, manager of Network Associate's McAfee Labs. A number of antivirus vendors sell products that block hostile Java applets and ActiveX controls.

ABOUT ANTIVIRUS TECHNOLOGY

Several basic antivirus technologies are used in a variety of products and research projects, including those based on biological models.

Scan-based virus detection

Many antivirus programs scan a system for infected files, either at a preset time (such as system startup) or after certain events (such as a file download).

Many scanning techniques search system memory, boot records, boot sectors, and files for byte strings that match the strings of known viruses stored in a lookup table. To maintain reliability, users must update lookup tables frequently. Viral scanning at the server level or via a LAN manager can help protect against infection being spread throughout a network.

Other scanning techniques, including the biologically based ones developed by IBM and the University of New Mexico, can also detect unknown viruses.

Behavior-based virus detection

Some antivirus programs monitor a computer system for the type of behavior that a virus causes, such as a file trying to duplicate itself or multiple programs launching at the same time. However, these systems will also give false alarms if legitimate activities produce such behavior, said McAfee's Talbott.

McAfee's products use pattern-matching algorithms to analyze a computer system's behavior.

Symantec's Merritt said Norton Antivirus 4.0 uses a proprietary technology called Bloodhound to look for specific anomalous behaviors. Merritt said Norton Antivirus 4.0 can also execute a suspicious file in a virtual environment shielded from the main system, such as a Word for Windows emulator, to determine if it has a virus infection.

Comparing behavior-based and biologically based antivirus systems, Talbott said, "Basically, it all comes down to heuristics, which all of us in the antivirus community are looking at and implementing."

However, IBM's Kephart said the biological model offers special opportunities to take advantage of the similarities between computer and human viruses.

The University of New Mexico's Forrest agreed and said the key question is "What is it about the [human] immune system that is really important to computing?" Both Forrest and Kephart said it is important to adopt what is useful and discard what is not.

Therefore, to get the best results as research continues into biologically based computer antivirus technology, Forrest said, "We need a rich dialogue between biology and computation." ❖

Lee Garber is a staff editor at Computer *magazine. Contact him at l.garber@computer.org.*

Richard Raucci is a freelance technology writer based in San Francisco. Contact him at rraucci@well.com.