Science
AAAS

Technological Networks and the Spread of Computer Viruses

Justin Balthrop *et al.*

Science **304**, 527 (2004);
DOI: 10.1126/science.1095845

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2004 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

process? Many MHC class I–binding peptides have been isolated and sequenced, and no spliced peptides have previously been described. One could therefore argue that splicing is uncommon. Alternatively, because the peptides identified chemically are the dominant peptides from the hundreds bound to the MHC class I molecules of a cell, the process may be common but inefficient. Cytotoxic CD8 T cells are famously sensitive, detecting very small numbers of MHC class I–peptide complexes on the surface of APCs. In the most extreme example, Sykulev *et al*. (*5*) even suggest that a single MHC class I–peptide complex could be detected. Examples of peptides inefficiently generated but recognized by CD8 T cells abound in the literature. Some come from proteins derived from alternate reading frames, 5′- or 3′-untranslated sequences, or even introns [reviewed in (*1*)]. A recent example was produced by aberrant initiation of translation at a leucine codon instead of the normal methionine codon (*6*).

It is telling that the two spliced peptides cited here are derived not from foreign proteins but from normal human proteins. This is commonly the case for peptides recognized by tumor-specific CD8 T cells. MHC class I molecules bound to rare self peptides that do not induce immunological tolerance probably constitute the only antigenic complexes normally available to CD8 T cells for tumor recognition. The CD8 T cells activated by such complexes potentially could react with normal tissues, which raises the interesting possibility that spliced peptides might occasionally be the targets of CD8 T cells in autoimmune diseases.

References

1. N. Shastri, S. Schwab, T. Serwold, *Annu. Rev. Immunol.* **20**, 463 (2002).
2. K.-I. Hanada, J. W. Yewdell, J. C. Yang, *Nature* **427**, 252 (2004).
3. N. Vigneron *et al.*, *Science* **304**, 587 (2004); published online 17 March 2004 (10.1126/science.1095522).
4. J. Lowe *et al.*, *Science* **268**, 533 (1995).
5. Y. Sykulev, M. Joo, I. Vturina, T. J. Tsomides, H. Eisen, *Immunity* **4**, 565 (1996).
6. S. R. Schwab, K. C. Li, C. Kang, N. Shastri, *Science* **301**, 1367 (2003).

COMPUTER SCIENCE

Technological Networks and the Spread of Computer Viruses

Justin Balthrop, Stephanie Forrest, M. E. J. Newman, Matthew M. Williamson

Computer viruses and worms are an increasing problem throughout the world. By some estimates 2003 was the worst year yet: Viruses halted or hindered operations at numerous businesses and other organizations, disrupted cash-dispensing machines, delayed airline flights, and even affected emergency call centers. The Sobig virus alone is said to have caused more than $30 billion in damage (*1*). And most experts agree that the damage could easily have been much worse. For example, Staniford *et al.* describe a worm that could infect the entire Internet in about 30 s (*2*). A worm of this scale and speed could bring the entire network to a halt, or worse.

The term virus refers to malicious software that requires help from computer users to spread to other computers. E-mail viruses, for instance, require someone to read an e-mail message or open an attached file in order to spread. The term worm refers to infections that spread without user intervention. Because they spread unaided, worms can often spread much faster than viruses. Computer infections such as viruses and worms spread over networks of contacts between computers, with different types of networks being exploited by different types of infections. The structure of contact networks affects the rate and extent of spreading of computer infections, just as it does for human diseases (*3*–*7*); understanding this structure is thus a key element in the control of infection.

Both traditional and network-based epidemiological models have been applied to computer contagion (*3*–*5*). Recent work has emphasized the effects of a network's degree distribution. A network consists of nodes or vertices connected by lines or edges, and the number of edges connected to a vertex is called its degree. Of particular interest are scale-free networks, in which the degree distribution follows a power law, where the fraction p_k of vertices with degree k falls off with increasing k as $k^{-\alpha}$ for some constant α. This structure has been reported for several technological networks, including the Internet (*8*) and the World Wide Web (*9*, *10*).

Infections spreading over scale-free networks are highly resilient to control strategies based on randomly vaccinating or otherwise disabling vertices. This is bad news for traditional computer virus prevention efforts, which use roughly this strategy. On the other hand, targeted vaccination, in which one immunizes the highest degree vertices, can be very effective (*11*, *12*). These results rely crucially on the assumption that the degree distribution follows a power law, and also that the contact pattern is static.

Many technological networks relevant to the spread of viruses, however, are not scale-free. Vaccination strategies focusing on highly connected network nodes are unlikely to be effective in such cases. Furthermore, network topology is not necessarily constant. In many cases the topology depends on the replication mechanism used by a virus and can be manipulated by virus writers to circumvent particular control strategies that we attempt. If, for instance, targeted vaccination strategies were found to be effective against viruses spreading over scale-free networks, viruses might be rewritten so as to change the structure of the network to some non–scale-free form instead.

To make these ideas more concrete, we consider four illustrative networks, each of which is vulnerable to attack: (A) the network of possible connections between computers using the Internet Protocol (IP), (B) a network of shared administrator accounts for desktop computers, (C) a network of e-mail address books, and (D) a network of e-mail messages passed between users.

In network A, each computer has a 32-bit IP address and there is a routing infrastructure that supports communication between any two addresses. We consider the network in which the nodes are IP addresses and two nodes are connected if communication is possible between the corresponding computers. Many epidemics spread over such an IP network. Notable examples include the Nimda and SQLSlammer worms.

Network B is a product of the common operating system feature that allows computer system administrators to read and write data on the disks of networked ma-

J. Balthrop and S. Forrest are in the Department of Computer Science, University of New Mexico, Albuquerque, NM 87131, USA; S. Forrest is also at the Santa Fe Institute, Santa Fe, NM 87505, USA. M. E. J. Newman is in the Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA. M. M. Williamson is at HP Laboratories Bristol, Filton Road, Stoke Gifford, Bristol BS34 8QZ, UK. E-mail: forrest @cs.unm.edu

chines. Some worms, including Nimda and Bugbear, can spread by copying themselves from disk to disk over this network.

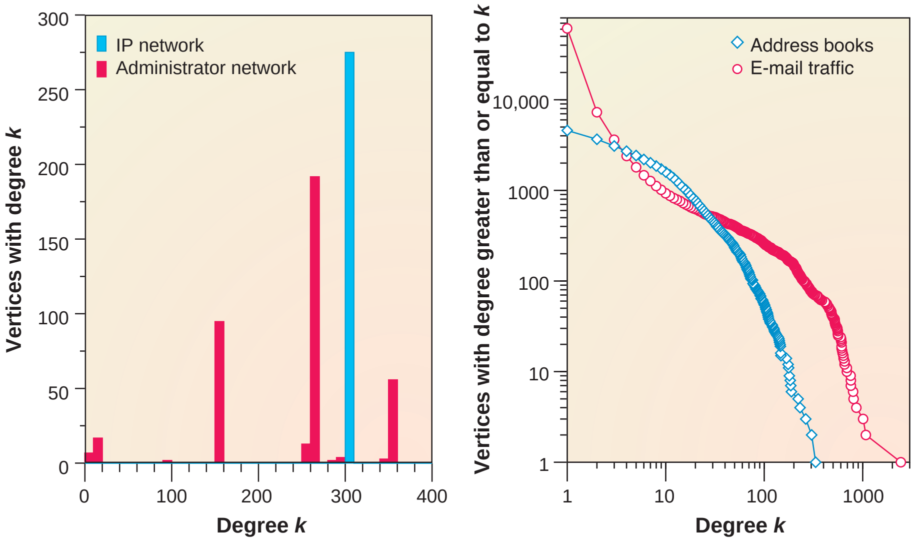
Network C has nodes representing users and a connection from user *i* to user *j* if *j*'s e-mail address appears in *i*'s address book. Many e-mail viruses use address books to spread (for example, ILoveYou). A closely related network is network D, in which the nodes represent computer users, and two users are connected if they have recently exchanged e-mail. Viruses such as Klez spread over this network.

Degree distributions have been measured for examples of each of the four networks (see the figure). In network A, all vertices have the same degree, so the distribution has a single peak at this value (blue histogram). In network B, the distribution consists of four discrete peaks, presumably corresponding to different classes of computers, administrators, or administration strategies (red histogram).

The two e-mail networks have more continuous distributions and are shown as cumulative histograms. Although neither network has a power-law degree distribution, both have moderately long tails, which suggests that targeted vaccination strategies might be effective. However, calculations show that for network C, about 10% of the highest degree nodes would need to be vaccinated to prevent an epidemic from spreading (*7*), whereas network D would require about 80%. The first of these figures is probably too high for an effective targeted vaccination strategy, and the second is clearly far too high. (Targeted vaccination would be entirely ineffective in the other two networks as well, because the nodes are much more highly connected.)

The two e-mail networks illustrate the ways in which different virus replication strategies can lead to different network topologies. An e-mail virus could look for addresses in address books, thereby spreading over a network with a topology like that of network C, or it could search through other files or folders on the machine for addresses of senders and recipients of archived e-mail messages, giving a topology more like network D. Another example is provided by the Nimda virus, which infects Web servers by targeting random IP addresses, producing a network like network A. However, if the virus had a more intelligent way of selecting IP addresses to attack (e.g., by inspecting hyperlinks), then it might spread over a topology more like that of the Web, which is believed to have a power-law degree distribution (*9*, *10*).

A control strategy is needed that is immune to changes in network topology and that does not require us to know the mechanisms of infections before an outbreak. A number of methods have been proposed (*13*). One such strategy is throttling, first introduced for the control of misbehaving programs (*14*) and recently extended to computer network connections (*15*). In this context, throttling limits the number of new connections a computer can make to other machines in a given time period. Because it works by limiting spreading rates rather than stopping spread altogether, the method does not completely eliminate infections but only slows them down. Frequently, however, this is all that is necessary to render a virus harmless or easily controllable by other means (*16*).



Infectious connections. (Left) Degree distributions for the IP (blue) and administrator (red) networks. The administrator network data were collected on a system of 518 users of 382 machines at a large corporation. Computers in this network are connected if any user has administrative privileges on both computers. **(Right)** Cumulative degree distributions for the e-mail address book (blue) and e-mail traffic (red) networks. The e-mail address book data were collected from a large university (*7*). The e-mail traffic data were collected for complete e-mail activity of a large corporate department over a 4-month period (*18*).

Throttling is most effective when viruses generate traffic at a rate significantly higher than normal network communications. Luckily this is true for many common protocols and the viruses that exploit them (*15*, *17*). For a virus to spread, it needs to propagate itself to many different machines; to spread quickly, it must do so at a high rate. For example, the Nimda worm attempts to infect Web servers at a rate of around 400 new machines per second, which greatly exceeds the normal rate of connections to new Web servers of about one per second or slower (*15*).

A throttling mechanism that limited connections to new Web servers to about one per second would slow Nimda by a factor of 400 without affecting typical legitimate traffic. This could easily be enough to change a serious infection into a minor annoyance, which could then be eliminated by traditional means. Slowing the spread of Nimda by a factor of 400 (from a day to more than a year) would have allowed plenty of time to develop and deploy signatures and prophylactic software patches. (Of course, if throttling were implemented on only a subset of the nodes in a network, then infections could spread more easily.) In addition to reducing virus spread, throttling has the practical benefit of reducing the amount of traffic generated by an epidemic, thus reducing demand on networking equipment—often the primary symptom of an attack.

Targeted vaccination strategies for the control of computer viruses are unlikely to be generally effective because the networks over which viruses spread are not sufficiently dominated by highly connected nodes, and because network topology can be influenced strongly by the way in which a virus is written. Throttling provides a promising alternative strategy that works with any network topology and can greatly reduce viruses' impact by slowing their spread to the point where they can be treated by conventional means. The disparity between the speed of computer attacks (machine and network speed) and the speed of manual response (human speed) has increased in recent years. If this trend continues, automated mechanisms like throttling will likely become an essential tool, complementing the largely manual approach of software patching in use today. The idea of rate limits is not specific to viruses, and could be applied to many situations in which an attack or cascading failure occurs faster than possible human response.

References and Notes

1. Citations documenting these events are listed on www.cs.unm.edu/~judd/virus.html, as are citations to each of the viruses and worms mentioned above.
2. S. Staniford, V. Paxson, N. Weaver, in *Proceedings of the USENIX Security Symposium* (USENIX Association, Berkeley, CA, 2002), pp. 149–167.
3. J. O. Kephart, S. R. White, in *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy* (IEEE Computer Society, Los Alamitos, CA, 1991), pp. 343–359.
4. R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
5. A. L. Lloyd, R. M. May, *Science* **292**, 1316 (2001).
6. H. Ebel, L.-I. Mielsch, S. Bornholdt, *Phys. Rev. E* **66**, 035103 (2002).
7. M. E. J. Newman, S. Forrest, J. Balthrop, *Phys. Rev. E* **66**, 035101 (2002).
8. M. Faloutsos, P. Faloutsos, C. Faloutsos, *Comput. Commun. Rev.* **29**, 251 (1999).
9. R. Albert, H. Jeong, A.-L. Barabási, *Nature* **401**, 130 (1999).
10. J. M. Kleinberg, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, in *Proceedings of the International Conference on Combinatorics and Computing*, vol. 1627 of *Lecture Notes in Computer Science* (Springer, Berlin, 1999), pp. 1–18.
11. R. Albert, H. Jeong, A.-L. Barabási, *Nature* **406**, 378 (2000).
12. D. S. Callaway, M. E. J. Newman, S. H. Strogatz, D. J. Watts, *Phys. Rev. Lett.* **85**, 5468 (2000).
13. D. Moore, C. Shannon, G. Voelker, S. Savage, in *Proceedings of the 22nd Annual Joint Conference of IEEE Computer and Communication Societies (INFOCOM)* (IEEE Communications Society, New York, 2003), pp. 1901–1910.
14. A. Somayaji, S. Forrest, in *Proceedings of the 9th USENIX Security Symposium* (USENIX Association, Berkeley, CA, 2000), pp. 185–197.
15. M. M. Williamson, in *Proceedings of ACSAC Security Conference* (IEEE Computer Society, Los Alamitos, CA, 2002), pp. 61–68.
16. M. M. Williamson, *Complexity* **9**, 34 (November–December 2003).
17. M. M. Williamson, in *Proceedings of ACSAC Security Conference* (IEEE Computer Society, Los Alamitos, CA, 2003), pp. 76–85.
18. J. R. Tyler, D. M. Wilkinson, B. A. Huberman, in *Communities and Technologies*, M. Huysman, E. Wenger, V. Wulf, Eds. (Kluwer, Dordrecht, Netherlands, 2003), pp. 81–95.

19. We thank J. Gassaway for help collecting the e-mail address book data set, C. Hickman for the administrator data set, and J. Tyler and B. Huberman for the e-mail traffic data set. Supported by the James S. McDonnell Foundation, NSF, Defense Advanced Research Projects Agency, Intel Corp., and Santa Fe Institute.

IMMUNOLOGY

CD8αα and T Cell Memory

Sangwon V. Kim and Richard A. Flavell

Our immune systems have the ability to remember past infections and to respond robustly upon subsequent infection with the same pathogen. This immunological memory provides long-term protection and makes vaccination possible. Long-term protection is mediated by memory T and B cells as well as by effector B cells (plasma cells that produce antibody), and depends on maintenance of these cell populations. In the case of CD8 memory T cells, preservation depends on the cytokines interleukin (IL)–7 and IL-15 (*1*). But, after the initial encounter with antigen, how are memory cells of the immune system produced? It is well established that in response to an invading pathogen, T cells specific for foreign antigen expand exponentially and differentiate into effector cells that clear the infection. At the end of the immune response, most of the expanded T cells die with a small portion of the survivors remaining as memory T cells. Two questions continue to intrigue immunologists (*2*). First, how do memory T cell precursors survive when the majority of their brethren die off? Second, how do T cells make a decision about their fate: to become either short-lived effector cells or long-lived memory cells? Answers remain elusive, as it is difficult to identify those T cells destined to become memory cells among the activated T cell population. Enter Madakamutil *et al*. (*3*) on page 590 of this issue, who shed light on this puzzle by showing that the homotypic form of CD8, CD8αα, is selectively expressed by CD8 memory T cell precursors and is required for their survival.

What prompted the authors to look at CD8αα expression in memory T cells? The same group originally developed tetramers of thymic leukemia antigen (TL) in an effort to discover the natural TL ligand (*4*). TL is a nonclassical major histocompatibility complex class I molecule that is abundantly expressed by intestinal epithelial cells. It does not present antigen, but instead binds to CD8αα on intraepithelial lymphocytes (IELs), modifying their response through activation of T cell receptors (*4*). Because IELs have a memory T cell phenotype (*5*), Madakamutil *et al.* explored whether CD8αα is important for the formation of CD8 memory T cells. They found that only a portion of CD8 T cells activated during the primary immune response expressed CD8αα; in contrast, CD8αα was expressed by most CD8 T cells involved in the recall immune response. Interestingly, upon anti-CD3ε stimulation of spleen cells (splenocytes) in vitro, CD8αα-negative splenocytes underwent apoptosis whereas CD8αα-positive splenocytes did not. This finding implies that CD8αα-positive T cells represent memory precursors that evade cell death during the contraction phase of the primary response.

To see whether CD8αα-positive T cells are really precursors of CD8 memory cells, Madakamutil *et al*. sorted mouse CD8 effector T cells into CD8αα-positive and CD8αα-negative populations according to their response to lymphocytic choriomeningitis virus. They then transferred the two populations into naïve mouse recipients. When the recipient mice were rechallenged with the virus at day 40 after transfer, the animals that received CD8αα-positive cells showed a robust memory response as judged by the frequency of specific T cells producing interferon-γ (IFN-γ). In contrast, mice that received CD8αα-negative T cells did not show a recall response to the virus.

Does CD8αα really play an active role in the generation of CD8 memory cells, or does its expression merely correlate with the formation of memory T cell precursors? To answer this question, Madakamutil *et al*. (*3*) used CD8 enhancer ($E8_I$) knockout mice (*6*) that express the usual heterotypic CD8 molecule CD8αβ but do not express the homotypic form CD8αα. Intriguingly, $E8_I$ knockout mice displayed a frequency of antigen-specific, IFN-γ–producing CD8 T cells similar to that of wild-type mice at day 7 after viral infection. However, the number of antigen-specific memory T cells in $E8_I$ knockout mice decreased dramatically 50 days after primary challenge and also during rechallenge. This confirmed that CD8αα is required for the generation of memory T cells but not for short-lived effector cells.

How is CD8αα involved in the process of memory T cell generation? One possibility is that it up-regulates expression of the IL-2/IL-15 receptor β (Rβ) chain (also

S. V. Kim and R. A. Flavell are in the Section of Immunobiology, Yale University School of Medicine, New Haven, CT 06520, USA. R. A. Flavell is an Investigator of the Howard Hughes Medical Institute. E-mail: richard.flavell@yale.edu