The Code Red Epidemic: a Case Study

John Lamp
Faculty of Business and Law, Deakin University, Australia
Email: John.Lamp@deakin.edu.au

An analysis of log files from an immune World Wide Web server was used to discover the patterns of infection from the Code Red worm variants. Analogies are drawn to biological systems. The need for protection is commented on.

Keywords: EK10 Computer Viruses, EL08 IS Risk Management

INTRODUCTION

Computer viruses have become a fact of life. Once a rarity, the increasing use of networks and network applications, such as email, has resulted in their rapid dispersion after creation. Their increasing prevalence is providing data that can be analysed to obtain information and devise models on rates of infection, spatial dispersion and other aspects.

THE CODE RED WORM

The first variant of the Code Red worm (CODERED.A) was discovered on the 19th July 2001. Subsequently two variations of the worm were discovered, CODERED.B on 31st July 2001, and CODERED.C on 6th August 2001. The worm specifically attacks web servers running Microsoft's Internet Information Server (IIS) on Windows NT 4, 2000 or XP Beta platform. The worm exploits an unchecked buffer in the Internet Data Administration area of IIS, contained in idq.dll. It creates a buffer overflow and gains complete control over the server. The actions the worm then takes is dependent on the variant of the worm.

Variant A is relatively benign. If the system date is before the 20th of the month, it generates random IP addresses and sends copies of itself to other web servers. If the system date is between the 20th and 28th, it executes a distributed denial of service (DDoS) attack on 198.137.240.91 (www1.whitehouse.gov). This server has now been moved to a different IP address. It also checks for a file called C:\NOTWORM. If this file exists, the worm goes dormant. It also causes the server on which it resides to display the message "Welcome to http://www.worm.com! Hacked By Chinese!" when accessed. Typically you see something like the following line in the log file, where [240*"N"] is a string of 240 "N"s.

```
GET /default.ida?[240*"N"]%u9090%u6858%ucbd3%u7801%u9090
```

Variant B fixes a bug in variant A, which caused variant A to generate the same random addresses. It displays the hacked message only if the system language is not English.

Variant C can be identified in the log files as it uses a string of "X"s rather than "N"s. Variant

Manuscript received: August 2001
Editor: John Roddick

C seems to work properly only on Windows 2000 machines. It causes NT 4 machines to crash. Variant C spawns either 300 or 600 threads to infect other machines compared with the earlier variants' 99 threads. Variant C is also more malicious in that it additionally installs a trojan version of Explorer and disables system file protection by modifying the registry. It also maps the C: and D: drive to web accessible directories /C and /D, thus making any file on the machine web accessible. Yet another good reason to have directory browsing turned off as default! It does not have a dormant period, nor undertake a DDoS attack. (Trend, 2001a; Trend, 2001b; Trend, 2001c; eEye, 2001; Microsoft, 2001)

LOG FILE ANALYSIS

The machine that supplied the case study data is running the Sambar Web Server (Sambar, 2001) on a Windows 2000 Professional platform. It is not behind any firewall. The log file accumulated has the format in Table 1.

remotehost	Remote hostname or IP address number if DNS is not enabled/available.
rfc931	The remote login name of the user. (This is not implemented by the Sambar Server).
authuser	The username of the authenticated user. This is available when using password protected WWW pages.
[date]	Date and time of the request.
"request"	The HTTP request line as it came from the client.
status	The HTTP response code returned to the client. Indicates whether or not the file was successfully retrieved, and if not, what error message was returned.
bytes	The number of bytes transferred. If the status is 200 and bytes are 0, the dynamic page size could not be determined.
msec	The time in milliseconds that it took the server to respond to the request.
"referer"	The url the client was on before requesting this url.
"agent"	The browser the client is using.

Table 1: Sambar Server log file format (Sambar, 2001)

A CODERED.A infection attempt was first logged on the case study machine on 20 July 2001. It later re-appeared on 30 July 2001, following which it has become a permanent feature, with CODERED.C dominant at around 150 attempts per day. In total 2,539 incidents have been recorded up to and including 20th August 2001. The grep utility was used to extract log file lines caused by virus infection attempts. These records were imported into an Access database and reformatted so that a cross tabulation of date against variant of virus could be run. The results were then graphed using Excel. (fig 1)

The log file for 20th of August 2001 has no entries for CODERED.A or .B, indicating it has entered its dormant period. CODERED.C is still being logged. Its trend is downward, but slower than that of variant A/B.

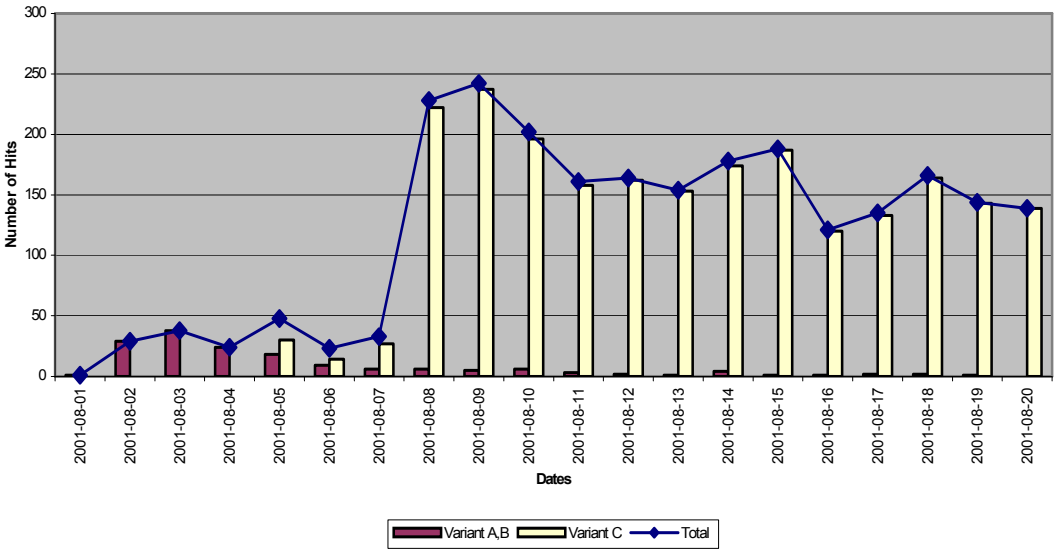

Figure 1: Code Red Infection Attempt Pattern – August

BIOLOGICAL ANALOGUES

Information technology is an area rich in metaphor, and the terms used to describe biological epidemics are often used loosely to describe the actions of computer viruses. (eg Kephart *et al*, 1993, 1997) In particular, epidemiological terms, such as *infectives* that introduce the disease, *susceptibles* that are able to be infected, and *immunes* that cannot be infected, are often used. From the quantitative theory behind biological epidemic analysis, graphs may be constructed, which describe the progress of the epidemic. (Poole, 1974) (fig 2

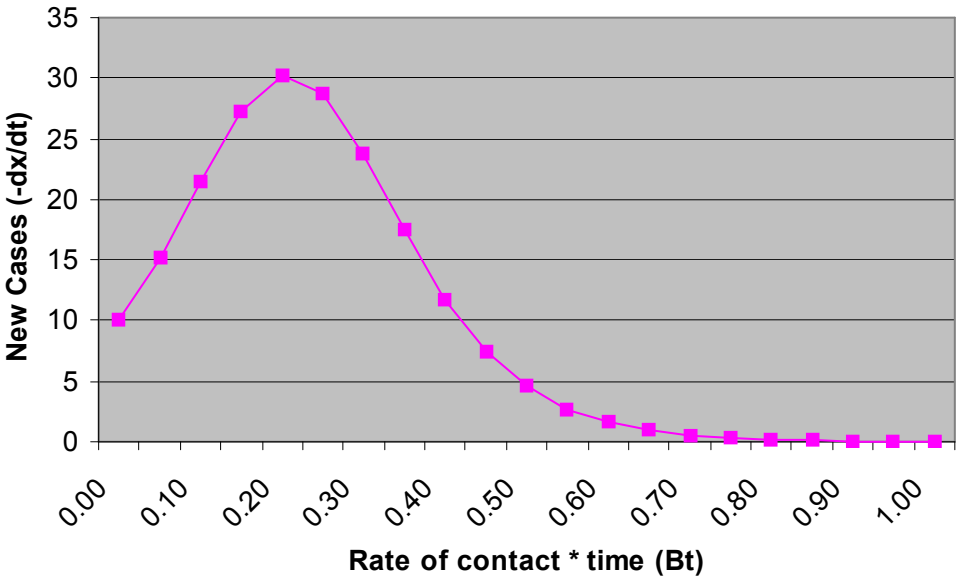

Figure 2: Deterministic epidemic curve for a simple epidemic (after Bailey, 1957)

Biological epidemic analysis is based on the number of new cases reported, and the requirement to report is, most often, a statutory one placed on the health profession. In this analysis, what was actually recorded was unsuccessful infection attempts. There is no parallel to this in classical epidemiology – it is not really possible to say, "That breath contained a flu virus particle that did not infect me!" Intuitively it might seem that the number of infection attempts should be related to the number of infectives in the population, but this requires a number of assumptions on how infectives are actually attempting to spread their infection, and needs closer examination. Certainly the literature in this area has a large number of mathematical models, which could be useful in understanding and predicting the impact of computer viruses on the normal operation of systems. Equally, there are a number of areas, such as modelling the spatial spread of biological epidemics, which have not been able to be undertaken effectively, but may be able to be addressed in the information technology domain. These could then be examined in the biological domain.

CONCLUSION

The increased prevalence of viruses and the ability to log attempts by viruses to attach machines is resulting in a greater amount of data, which could form the basis of new methods for modelling their distribution. Existing models of biological epidemics may be applicable to epidemics of computer viruses. In this specific case, large numbers of infection attempts are still being logged despite a corrective patch being issued on 18th June 2001, prior to the discovery of the virus in the wild. (Microsoft, 2001)

REFERENCES

Bailey N J T (1957) *The Mathematical Theory of Epidemics*, Charles Griffin & Co, London

eEye Digital Security (2001) *AD20010618 All versions of Microsoft Internet Information Services Remote buffer overflow (SYSTEM Level Access)* [Online]
Available: http://www.eeye.com/html/Research/Advisories/AD20010618.html
Last Accessed: 2001-08-20

Kephart J O., D M. Chess, and S R. White (1993) "Computers and Epidemiology" *IEEE Spectrum* [Online]
Available: http://www.research.ibm.com/antivirus/SciPapers/Kephart/Spectrum/Spectrum.html
Last Accessed: 2001-08-20

Kephart J O., G B. Sorkin, D M. Chess and S R. White (1997) "Fighting Computer Viruses" *Scientific American* [Online] Available: http://www.sciam.com/1197issue/1197kephart.html Last Accessed: 2001-08 20

Microsoft (2001) *MS01-033 Unchecked Buffer in Index Server ISAPI Extension Could Enable Web Server Compromise* [Online]
Available: http://www.microsoft.com/technet/treeview/default.asp?url=/technet/security/bulletin/MS01-033.asp
Last Accessed: 2001-08-20

Poole R (1974) *An Introduction to Quantitative Ecology*, McGraw-Hill, New York

Sambar T (2001) *Sambar Server* [Online] Available: http://www.sambar.com/ Last Accessed: 2001-08-20

Trend (2001a) *CODERED.A* [Online] Available:
http://www.antivirus.com/vinfo/virusencyclo/default5.asp?VName=CODERED.A Last Accessed: 2001-08 20

Trend (2001b) *CODERED.B* [Online] Available:
http://www.antivirus.com/vinfo/virusencyclo/default5.asp?VName=CODERED.B Last Accessed: 2001-08-20

Trend (2001c) *CODERED.C* [Online] Available:
http://www.antivirus.com/vinfo/virusencyclo/default5.asp?VName=CODERED.C Last Accessed: 2001-08-20

BIOGRAPHICAL NOTE

John Lamp is a lecturer in the School of Management Information Systems at the Waurn Ponds campus of Deakin University. He originally graduated as a zoologist from the University of Tasmania. He worked in the public sector in a variety of programme and support positions covering information technology, change management and strategic management. He commenced an academic career at the University of Tasmania in 1995. In 1998 he joined Deakin University. John's interests include systems development methodologies, and project management.