Application of Models from Epidemiology to Metrics for Computer Virus Risk -- A Brief Update

Joan L. Aron and Ronald A. Gove
Science Communication Studies, Columbia, Maryland, USA and Science Applications International Corporation, McLean, Virginia, USA

Keywords: Mathematical model, risk assessment, computer virus, malicious code, computer simulation, information security metric, computer virus survey.

1. INTRODUCTION

This brief update describes research on the maintenance of integrity in information systems via the establishment of an organisational environment that will prevent the damage caused by external agents. The main focus is on the development of information security metrics and a computer simulation model for the threat of computer viruses in organisations. Early results from this research project were presented at the IFIP TC11 Working Group 11.5's Second Working Conference on Integrity and Internal Control in Information Systems (Aron and Gove, 1998). This brief update summarizes subsequent work conducted by a team comprising Science Applications International Corporation, Science Communication Studies and the Towson University Applied Mathematics Laboratory (SAIC 1999a, 1999b).

2. EXTENDED ABSTRACT

In the case of computer viruses and some other types of malicious code, the formal analogies to the infection dynamics of biological viruses permit the utilisation of epidemiological concepts in the development of metrics. The first phase of the project demonstrated how a simple epidemiological model of computer viruses provides insights about the importance of several metrics:

1. frequency of contact;
2. utilisation of antiviral software;
3. effectiveness of antiviral software;
4. likelihood of notifying other people about computer viruses detected;
5. frequency of updating antiviral software.

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35501-6_14

M. E. van Biene-Hershey et al. (eds.), *Integrity and Internal Control in Information Systems*

In subsequent work, epidemiological concepts from environmental health have added a new perspective that permits better integration of infection control and organisational information security metrics (Beaglehole *et al.*, 1993). Information security metrics for an organisation are analogous to environmental health indicators, which are defined as an expression of the link between environment and health, targeted at an issue of specific policy or management concern and presented in a form that facilitates interpretation for effective decision-making (Briggs *et al.*, 1996). The benefits of a broader analysis of an organisation's environment are: (1) an understanding of how organisational complexities influence the risk associated with computer viruses; (2) a representation of organisational factors in terms familiar to those making decisions; and (3) integration of assessment of risk from multiple threats.

The earlier analysis was enhanced by improving the quality of available data on computer virus infections in organisations and by using those data to increase the sophistication of the representation of actual computing environments. Specifically, the aims were to:

1. collect real-world data on virus incidents to validate the suggested metrics and to identify, if possible, other risk factors;
2. analyse the collected data using statistical analysis techniques commonly used in epidemiology and derive some potential risk factors;
3. implement a more comprehensive simulation model for examining the results of modifying risk parameters for virus incidents.

A computer virus epidemiology survey was set up for anonymous access via browser on the World Wide Web. A list of possible responses in the form of radio buttons or checkboxes was provided for each survey question. Radio buttons permit the selection of only one item in a list, while checkboxes permit the selection of more than one item. The survey, which is located at URL http://survey.secureweb.net, has been online (subject to some brief downtime) from June 1998 to June 1999, although only the 48 entries obtained from respondents by January 25, 1999 are discussed here. The survey has six sections that focus on:

1. basic demographic information about the organisation and the workgroup of the respondent;
2. views of the respondent about the threat of computer viruses and general aspects of protection;
3. computer virus experience over the past twelve months in the respondent's workgroup;
4. system environment including groupware, network management and organisational turnover;
5. practices for sharing files, ranging from physical media to shared network volumes and internet services;
6. practices for system protection, ranging from URL filtering, usage of antiviral software, reporting of computer viruses to system administrators to user training about information security.

A review of the quality of the data resulted in the removal of entries from six respondents who submitted the survey despite the fact that they stopped answering questions midway through the survey. (The website maintains no records of those who started the survey but did not submit it.) The answers that were provided by the remaining respondents about recent computer virus experience appeared to be internally consistent. A number of questions had a high percentage (> 30%) of non-response (the respondent selected "don't know" or skipped the question), suggesting that some questions would have been more useful in a different form or should have been omitted altogether.

Potential risk factors in an organisational environment were analysed by examining associations between characteristics of the organisational environment and computer virus incidents in the form of a 2x2 table, i.e., a cross-tabulation of two qualitative (categorical) variables at two levels. One variable measures the presence or absence of a particular computer virus experience and the other measures the presence or absence of a particular potential risk factor in an organisational environment. For a variable with multiple levels of response, such as different frequencies of computer virus incidents, it was necessary to create a new variable combining levels into just two levels. There were not enough observations in the study for the application of multivariate analysis, which examines the joint effect of multiple variables.

Two different statistics were used for identifying possible relationships: the p-value of a chi-squared test of the null hypothesis and the odds ratio (OR). The chi-squared test examines the probability that the observed differences could appear by chance under the (null) hypothesis that there is no association between the risk factor and the computer virus experience. A smaller p-value indicates greater significance, i.e., that an apparent association is less likely to be the result of chance alone. P-values that are less than or equal to 10% pass an initial screening. Larger sample sizes and smaller p-values are needed for confirmation. The OR is a statistic of the magnitude of association, similar to a risk ratio. The larger the OR (e.g., 2 for doubling risk or 3 for tripling risk), the larger the association. Significance testing of the OR is also possible with larger sample sizes.

It is possible for the OR to be less than 1, i.e., the factor in question is protective. In this situation, it is especially critical to understand the original variables and how the measured risk should be interpreted. Also, some factors may show an association with risk even if common sense would dictate that they are not risk factors. For example, if people adopt a protective behaviour after a computer virus incident has occurred, then a protective behaviour may be associated with risk of computer virus incidents in a cross-sectional study (which means that responses to the survey measure one point in time). Another problem is that, with so many variables tested, some associations may be spurious because of the role of statistical chance. Thus it is important to remember that an association does not necessarily mean causation.

In parallel with the acquisition of new data, a new simulation model was developed to handle more complex characterisations of the organisational environment. The

simulation development environment is object-oriented MODSIM, a CACI product at the Towson University Applied Mathematics Laboratory. With object-oriented programming, it is possible to represent each virus on each computer instead of aggregated numbers of infected computers, allowing more detail on types of computer viruses and strategies for control. MODSIM also has rich capabilities for using random number generators to construct stochastic variables.

The survey data show, as expected, that risky behaviours include using groupware, sharing physical media every day, sending or receiving file attachments every day, and uploading and downloading files over the Internet every day. Sharing computers and not reporting virus infections in the computers of others are risks for severe computer virus incidents. Not using groupware is protective in general, as are central anti-virus policies and environments in which most people report computer viruses to network management.

Besides general risk factors analysed across all organisations, interactions between different characteristics are important. The survey data show that the strength of several risk factors depends on the size of the individual's workgroup or the size of the individual's organisation as a whole. In this analysis, a large organisation has over 1,000 enduser workstations and a large workgroup has over 100 enduser workstations. Small workgroups/organisations tend to be less likely to use groupware, more likely to share computers, less likely to report viruses and less likely to have a central anti-virus policy. Large workgroups/organisations, on the other hand, are more likely to use groupware, less likely to share computers, more likely to report viruses and more likely to have a central anti-virus policy. In addition, some behaviours, such as the daily use of e-mail attachments, are indirectly associated with workgroup or organisation size through an association with other behaviours, such as the use of groupware. An association between groupware and sending e-mail attachments could account for the somewhat surprising observation that larger organisations/workgroups appear to be at greater risk for macro virus infections than their smaller counterparts despite the fact that larger enterprises are more likely to use better security practices, such as having and verifying compliance with a central anti-virus policy.

The simulation scenarios confirm earlier results that three important factors -- the use of antiviral software, fast detection in the absence of antiviral software and little exposure to outside networks -- are protective, but the scenarios also indicate some surprising interactions. Although the risk of propagating computer viruses increases with increased frequency of contacts, such as sending e-mail attachments, a high frequency of contacts can be protective for organisations. The protective aspect of communication appears to be related to an assumption that 90% of the user community is aware of the problem and will notify management if their antiviral software detects a computer virus. Whether or not a policy is protective or risky may depend on other organisational characteristics. The very definition of protective is complex because a policy may be effective when measured with different metrics of the impact of computer viruses. For example, adding antiviral software to an e-mail

server always reduced the number of days with any infected computers, but did not always reduce the number of distinct episodes (defined as periods of time during which computer virus infections are present every day). Assumptions about how detection leads to cleanup were probably the reasons for the counterintuitive exception that occurred in an environment where people were relatively slow to detect computer viruses in the absence of antiviral software.

The results suggest the need to examine the business culture and organisational behaviour to make recommendations for an organisation to reduce risk. The framework for analysis should include the use of different technologies, including emerging technologies, and associated risks. In this light, it is interesting to consider the recent explosive spread of computer viruses that attack software agents authorised to send electronic mail with attachments, a development anticipated years in advance (Kephart *et al.*, 1997). Since rapidly spreading viruses can cause considerable damage before updated antiviral software is distributed through the usual channels, approaches that are based solely on technological fixes may be inadequate and an understanding of protective behaviours in an organisation at risk may assume greater importance.

3. ACKNOWLEDGEMENTS

We appreciate the assistance of Cristina Schneider and Jon McKnight in the development of this project as well as the efforts of Mike O'Leary, Shiva Azadegan, and Shadi Alegheband at the Applied Mathematics Laboratory of Towson University, part of the University of Maryland System. This work was supported by the U.S. Department of Defense.

4. REFERENCES

1. Aron J.L. and Gove R.A. (1998) Application of Models from Epidemiology to Metrics for Computer Virus Risk, in Integrity and Internal Control in Information Systems. IFIP TC11 Working Group 11.5 Second Working Conference in Information Systems: Bridging Business Requirements and Research Results, Warrenton, Virginia, USA, November 19-20, 1998 (eds. S. Jajodia, W. List, G.W. McGregor and L.A.M. Strous), Kluwer Academic Publishers, Boston/Dordrecht/London, pp. 131-145.
2. Beaglehole R., Bonita R. and Kjellstrom T. (1993) Basic Epidemiology. World Health Organization, Geneva.
3. Briggs D., Corvalan C. and Nurminen M. (1996) Linkage Methods for Environment and Health Analysis. General Guidelines. World Health Organization, Geneva.
4. Kephart J.O., Sorkin G.B., Chess D.M. and White S.R. (1997) Fighting computer viruses. Scientific American. **277**(5), 88-93.

5. SAIC. (1999a) Final Report. Application of Models from Epidemiology to INFOSEC Assurance Metrics. January 29, 1999. Science Applications International Corporation, McLean, Virginia.
6. SAIC. (1999b) Further Analysis of the Computer Virus Epidemiology Survey. Addendum to the Final Report on Application of Models from Epidemiology to INFOSEC Assurance Metrics. February 26, 1999. Science Applications International Corporation, McLean, Virginia.