# Malware, Spam, and Mobile Security

© Stock.XCHNG

Both malicious software (malware) and spam are growing concerns for business, particularly as they now seem to be converging. Although 2006 has so far witnessed a decline in infected e-mail attachments, managed service provider Black Spider recently reported that spam can constitute 65% of corporate e-mail, and there are suggestions that up to 10% of spam now has the potential to trigger malware infection. Malware is also becoming more covert. In the last few years, worms (that spread over networks), and Trojans (that masquerade as legitimate software) have gained ground over viruses (that spread by infecting files). Most recently, Trojans have taken a significant lead.

From being dominated by arbitrary file destruction and attention-seeking, behaviour has also shifted towards suborning of victims to create networks of spam distribution robots, or theft of credentials in support of fraud. Typical of these trends is the Google tool bar Trojan that emerged in July this year. Advertised in a spam e-mail that linked to a download Web site, the Trojan resembled the Google tool bar but was, in reality, a "zombie client" that allowed remote control of the infected computer. There is also some evidence emerging of malware being used to encrypt or lock a user's files as a basis for attempted extortion.

The malware community has clearly matured, and is now primarily seeking financial gain, so consolidation of links between malware authors and organised crime is to be expected. For the future, we can therefore predict better presented and targeted spam, associated with a growing volume of increasingly diverse, well-disguised Trojans or fully clandestine malware more tightly focused on the exploitation of specific business sectors. This is a disturbing trend, but what troubles me most is a more general observation. Over the last few years, both US and UK national security surveys have consistently shown that although over 90% of corporate respondents deploy anti-virus tools, between 70% and 80% still experience malware infection. Clearly, a significant amount of malware is slipping under the radar despite the continuing well-directed efforts of anti-virus vendors, and irrespective of the changing specifics of infection vectors and malware behaviours.

Malware authors have gained an edge by responding ever more rapidly to vulnerability reports. In 2002, it took 46 days from the vulnerability announcement for the SLAPPER worm to appear. In 2004, SASSER took 17 days, and last year ZOTOB was released in only five. The term "zero day exploit" has been coined to describe this window of opportunity. However, ever-shortening development times mean we now face "zero hour" malware that constantly changes, seriously hindering the timely release and deployment of reactive countermeasures. At its peak in 2005, several variants of the MYTOB e-mail virus were released in a single day.

Particularly in the context of mobile working, this turnaround rate poses a growing problem in the face of continuing reliance on centrally managed, reactive anti-virus and spam management tools as the sole or primary line of defence. The tools are mature, but they still fundamentally rely on blacklists that require constant updating as new malware and spam are discovered (often on a daily basis). For that reason, they are at a disadvantage when attempting to protect an increasingly mobile workforce.

It is interesting to note the relative costs of addressing e-mail security at different points within the infrastructure. In the context of reactive countermeasures deployed within perimeter-oriented security architectures, the accepted wisdom is that it is cheaper to prevent spam from entering the organisation in the first place. This can be through appropriate security at the boundary, or indeed through the use of a managed security service, rather than reliance solely on controlling internal mail servers and client endpoints, and this remains basically correct for fixed desktop computing, particularly from the cost perspective. However, Mike Dalton of McAfee points out, "Employee negligence means there is a high risk of malware, viruses, worms, and Trojans being spread to the work network. It only takes seconds for an employee to attach an unprotected laptop or PDA to the work network and seriously expose the whole environment to infection." This argument is persuasive.

The growth in mobile working has fundamentally undermined the security of such architectures by blurring the boundaries of the corporate network. The gateway is no longer the sole real entry point to the network, and the population of connected devices is not short-term stable, so optimum deployment of reactive controls is increasingly problematic. Mobile computing devices can become infected via untrusted connections whilst detached from the corporate network, and thus whilst not protected by the corporate security infrastructure.

The resulting infection can then spread to the network when the mobile device is reconnected. Protection of mobile computing resources that rely on centralised update of reactive anti-malware measures will always tend to be behind schedule, providing a window of opportunity for infection. Once it gets a foothold, malware is now potentially capable of subverting or disabling countermeasures, so from the malware management perspective, it has become necessary to consider the corporate laptop as untrusted.

For the large corporation, developments in deperimeterised security architectures, as advocated by the Jericho Forum, offer promising if potentially costly solutions for the future. However, for the majority of smaller enterprises, and for all of us in the shorter term, more economic and mature solutions are necessary, particularly as the malware threat will inevitably be pushed downwards to those with smaller security budgets by improvements in security at the top. Considerable improvements in overall corporate protection can be made by applying anti-malware controls to the mobile user's laptop. However, due to the update lag, such controls should not merely be replicas of centralised reactive countermeasures. They must also include proactive, generic controls that do not require regular policing or updating. Whereas the aim of gateway anti-virus and spam management tools is to detect and prevent specific spam and malware entering the network or arriving on computing devices, the primary objective of workstation-based proactive controls should be to prevent any generic malware that may penetrate those defences from functioning.

Malware that cannot execute is not an immediate hazard, so its detection and elimination become a less urgent priority, and the anti-virus update lag becomes less critical. Preventing malware functioning generically is also a much simpler problem to manage than the detection of the latest variants, as the parameters of the problem are reasonably static. Although it is still not universally recognised, a significant level of first-line protection can be gained by appropriate configuration of the security options already built into operating systems, Web browsers, and e-mail clients. User privilege on the mobile workstation should be set to the minimum required for business purposes.

The generally excessively liberal security settings in Web browsers should be tightened. E-mail clients should be configured to prevent execution of embedded scripts and click-through of HTML links. A recent comment from the Virus Bulletin noted the enhanced robustness of Microsoft's new 64-bit Web browser as being primarily due to its inability to execute 32-bit Active-X. I believe this sends a clear signal about the security issues of active content. Such simple configuration changes, by preventing illicit execution of scripts and code, can provide a robust first line of defence.

However, for maximum protection, software execution management tools can prevent unauthorised or infected software running with improved granularity of control. Such tools offer protection by digitally signing approved applications, and preventing any program executing if it is either unsigned or does not match its predefined signature when it is invoked. Comparable signing tools are available to manage the authorisation of PDAs, memory sticks, and similar devices.

Supplementing centrally managed reactive spam and malware controls with basic workstation hardening has not, so far, received wide enough attention, particularly among SMEs, and in the lower echelons of the public sector. Nevertheless, such hardening is remarkably simple to deploy during equipment rollout. It offers significantly improved protection over reactive anti-malware solutions alone at relatively low capital outlay and minimal recurring cost. Bearing in mind the rate at which the malware threat landscape is now evolving, it also promises improved protection against future zero-hour threats by employing generic countermeasures that do not need constant updating in the face of short-term changes to the detail of threats.

**Mike Barwise**
An independent consultant specialising in information security strategy
m.barwise@bcs.org.uk