Securing documents proves tricky

While most users are by now aware of the potential risk for infection inherent with opening applications sent over the Internet, fewer are aware of the risks that can come from documents.

Seth Hardy, senior malware analyst for the threat research and response team at Symantec and part of MessageLabs prior to Symantec's purchase of the company, addressed the security challenges of documents in a session Wednesday at the SecTor Security Education Conference in Toronto.

"The problem with documents is that they greatly outnumber applications, and we just can't get rid of them – we need them to do business," Hardy said. "And in many cases, we're convinced that certain types of documents are safe."

Hardy focused on the two most major document types in the wild, those built around Object Linking and Embedding (OLE), and PDF files.

OLE files include most Microsoft Office documents and some versions of WordPerfect, as well as Windows Installer (.msi files). Although the risks of malicious macros have been well publicized, and many users are now taught to avoid running any unexpected macros, there are other risks as exploits can be taken advantage of to insert malicious code or run applications right in the document. Hardy said there have been seven attacks this year based on known OLE faults. Those can be defeated through patching, but therein lies the problem. Other challenges include the fact that scanning for these exploits can be very difficult, as in some cases only a few bytes separate the good from the bad.

With the macro-related challenges in Word documents, many organizations turned to Adobe's PDF as a way to avoid attacks. But while the PDF file may fare better than OLE documents in the security realm, it's hardly perfect. And it's getting worse.

Hardy cited the decision to allow the embedding of Flash content in PDF files in recent versions of Acrobat as "opening up a whole world of pain."

"The moment they started putting Flash into PDFs, something went horribly wrong." Hardy said. "Flash is in and of itself the most horrible thing in the history of ever, but not it's not only invading the Internet, but your documents as well."

Beyond the risk of rogue Flash applications running in PDF files, there's the ongoing challenge of PDF allowing JavaScript. The situation has been improved by the fact that the document now asks before it starts running invoked JavaScript, although Hardy questioned how many users read and think about such prompts before simply approving them.

And like some of the OLE attacks, the PDF exploits, particularly those using JavaScript, are particularly hard to pick up. In fact, Hardy said, signature-based antivirus simply doesn't cut it.

"They just don't work on this type of malware," he said. "You're just not going to catch it."

To help lock things down, security professionals have to "be able to validate everything" coming in via documents. On the OLE documents, he said to looke for embedded file and bad streams, and as a last resort, moving to the most recent XML-document-based versions of Microsoft Office help greatly, although

the XML documents have their own set of challenges.

While things aren't "as bad" on the PDF side, it can still be a tricky matter. Hardy urged security pros to parse all objects regardless of how they're used, and to look for JavaScript that an attacker is trying to disguise.

"If you find any obfuscated JavaScript in a PDF, you can throw it out right away," Hardy said. "Chances are, if you're finding hidden executable code in a document, something's wrong."