In 1982, a program known as “Elk Cloner” became the first computer virus to appear outside of a computer lab. It spread by floppy disk and did nothing more than alter a game to display a poem. Over the next two decades, viruses evolved into stealthy threats with the ability to compromise or destroy huge amounts of information. The growing popularity of the Internet provided viruses with an infection vector that could rapidly deliver malicious code to a worldwide audience. In the spring of 1999, the Melissa virus (a mass-mailing macro virus) illustrated this potential for great harm when it shut down Internet mail systems and caused millions of dollars of damage within a few hours of its discovery. Malicious software (“malware”) has grown even more sophisticated since then, with a number of highly complex viruses emerging to challenge traditional anti-virus methods. Our research explores a novel approach to detecting these new, advanced viruses.

As viruses have grown more complex, disk drive processors have grown more powerful. The typical modern SCSI disk has an underused 150-200 MHz processor with 8-16 MB of RAM. Our research explores the application of this processing power to virus detection. Disk processors are privy to the low-level actions of viruses that alter data on their hosts because all changes to data stored on a hard drive must first pass through the disk processor. We can therefore use the disk processor to observe disk requests and identify viruses based on the specific sequences of requests they make.

Current anti-virus software resides at the application level and relies heavily on a detection method known as string scanning. String scanning compares the bit patterns of binary files against a database of known virus signatures. Sophisticated “morphing” viruses evade string scanners by altering their own code structure between generations. The same effect is accomplished by authors of less-sophisticated viruses who hand-tailor existing code to create variants that no longer match the signature. Disk-level malware detection, however, is unaffected by these tricks because it looks for matches between disk traffic and viral behavior. This translates into a natural defense against morphing viruses: since the disk processor is judging intrinsic behavioral patterns, it cannot be fooled by surface code transformations.

Malware detection at the disk level is intended to work as a second line of defense against the most complex viruses. My research is focused on showing that viruses that evade operating system-level anti-virus software may be more simply caught at the disk level. To this end, I have been developing low-level behavioral signatures for viruses that cannot be easily or accurately identified by existing techniques. A Windows driver in a virtual machine simulates a plausible model of a future disk processor and provides me with logs of file system requests. I examine these logs to try to map specific disk requests to viral behavior, and then generalize this behavior into a pattern that describes an infection sequence.

The signatures I create need to be flexible enough to accommodate multiple generations of a morphing virus but still retain high precision and accuracy. The risk of false negatives (failing to identify an instance of a virus) drives a signature to be more general, while the risk of false positives (identifying benign software as a virus) leads to greater specificity. Either type of failure could have costly consequences for a consumer, so false positive and negative rates must be extremely small for a signature to be useful in practice. I need to carefully balance these and other issues when determining what behavior should be included in a signature.

The presentation will discuss the motivation for the research, explain my signature development process, and report on the accuracy and precision of the resulting signatures.