A number of highly sophisticated viruses have emerged in recent years to challenge traditional anti-virus methods. Foremost among these new anti-detection techniques are polymorphism and metamorphism, which automatically alter code structure between generations without changing the virus’s behavior. Polymorphic and metamorphic viruses are invisible to detectors that rely on string scanning because they have no static code. Similarly, virus authors often tweak existing viruses to create variants that are undetectable because they do not match the signature for the original virus. Since string scanning is the primary method of virus detection employed by anti-virus software, evasion of string scanning is a serious problem. Researchers have developed new techniques, the most effective of which is emulation; but even emulation is limited by its high computational cost and imperfection, and virus authors subsequently created anti-emulation techniques. Anti-virus software is therefore still currently unable to reliably detect advanced polymorphic and metamorphic viruses.

As viruses have grown more complex, disk drive processors have grown more powerful. The typical modern SCSI disk has a 150-200 MHz processor with 8-16 MB of RAM, and emerging standards such as OSD-2 are enriching communication between the host processor and disk. We propose that this processing power can be applied to a new method of disk-level virus detection. Because all changes to data stored on a hard drive must first pass through the disk processor, disk processors are privy to the low-level behavior of viruses that replicate by altering data on their hosts. We can therefore use the disk processor to observe disk requests and identify viruses based on the specific sequences of requests they make.

Virus detection at the disk level is particularly advantageous because of its low overhead, ability to detect morphing viruses, and immunity from layer-below attacks. Unlike emulation and related techniques, disk-level monitoring requires little extra effort from the CPU because it observes the behavior of currently running programs. Its defense against polymorphism and metamorphism comes naturally: a disk processor sees only behavioral patterns and cannot be fooled by sheer code complexity. Additionally, detection at the disk level cannot be undermined by malicious code running below the disk.

The immediate drawback to disk-level detection is that, unlike string scanning, behavior-based detection requires allowing a potential virus to run before it can be detected. However, this weakness is mitigated by the role of disk-level detection, intended precision of the behavioral signatures, and position of the disk drive processor. Many viruses can be caught adeptly with string scanning, and this research does not intend to supplant that. Instead, disk-level virus detection works as a complementary second line of defense: viruses that evade operating system-level software may be more simply caught at the disk level. Ideally, disk-level signatures will be exact enough to recognize a virus’s behavior before it has permanently damaged information or propagated. When the disk becomes suspicious of certain actions, it can silently trigger a backup of the blocks that are being written to. A compromised host would not be able to prevent the isolated disk from preserving the data.

In the process of creating precise disk-level signatures, we are also identifying what semantic information is needed. Currently, disk drives have no access to semantic information; they merely receive read or write requests to specific blocks. Some semantic information – e.g., process IDs, file types – may prove necessary for precise signatures. More expressive interfaces (such as the one provided by OSD-2) will offer a means to obtain some of this semantic information at the disk level. As we work to develop signatures, we will aim to maximize precision while minimizing semantic information. To experimentally create and test signatures, we are running a high-level driver in a Windows virtual machine and a low-level disk tracer in Linux and observing the behavior of viruses and benign applications.

We are particularly interested in viruses that cannot be easily or perfectly identified by string scanning or emulation. Our work addresses the need to balance early recognition, minimum semantic information, precision, and accuracy. The poster will present our research on disk-level behavioral signatures and report on our experiments with a prototype disk-level detector.