Lecture notes for “Reality of Malware Detection”

- Signature detection
  - Signature database, each entry a regex
  - Good sig: regex matches on >= 1 malware instance, 0 benign instances
    - [Explain regexes if needed]
  - Upsides
    - Match indicates something is malware (no false pos)
    - Good implementations are fast
  - Downsides
    - Limited to known instances
  - Errors
    - False positives: match on benign [McAfee once quarantined MS Office] [QA testing]
    - False negatives: malware authors revise malware to no longer match sig

- Known instance limitation
  - Now very troubling
  - Alteration of malware by author
    - [Think of today's software with version .A, .B, .C, .AA, .AB, .AC, etc]
    - Malware sig changes at speed of author
      - Changed sigs do not replace old sig in detector, always adding new entries, slows detector
  - Alteration of malware by malware
    - Malware changes itself before propagating
    - Sig changes at speed of propagation, which is **fast**
    - Signature detection cannot keep up, fails to detect
    - Types of alteration: polymorphism, metamorphism

- Polymorphism
  - [Def] Automatic self-modification
  - Syntactic
  - Often encryption with constantly changing key; only decryption loop remains exposed
    - Decrypt with just a couple of instructions, likely common code

- Metamorphism
  - [Def] Automatic self-reprogramming
  - Semantic
  - Use a different series of instructions that will end up causing the same effect
    - NOP insertion, register renaming, reordering independent instructions, insertion of control flow, opaque predicates, complete code regeneration
  - Detection: program equivalence is undecidable
  - Need to emulate execution and keep track of behaviors
    - Malware now looks for presence of virtual machines to defeat this

- End result: current malware detection is ~50-60% success

- Behavioral detection
  - Spyware
    - Sends data to a website; how is this abnormal?
    - Data comes from keyboard; how is this abnormal?
    - Data is confidential, flows from keyboard to malicious website; how is this detected?
  - Malware
    - Looks for presence of known monitors (VMs) and alters behavior