Factors for Differentiating Human from Automated Attacks

Kelly Greeling, Graduate Student at School of Information Science - UIUC
Alex Withers, Senior Security Engineer, National Center for Supercomputing Applications - UIUC
Masooda Bashir, Assistant Professor, School of Information Sciences - UIUC

• Background
  • Recent cyber-crime costs are at an all-time high and still skyrocketing.
  • Many Intrusion Detection Systems and Intrusion Protection Systems utilize behavior-based methodology, which seeks to identify a baseline for normal users that is then used to compare against real-time and non-real-time events in an effort to locate malicious activity.
  • The rise of automated attacks has created a great deal of noise for security personnel to wade through to identify malicious behavior and even with IDS systems, a human actor is still required to go through the logs to note is unusual activity is actually a threat.
  • If a human based attack is significantly different than an automated attack it would be extremely useful for security personnel to have a way to separate the behavior of an automated cyberattack tool from that of a human actor, as this would allow them to create separate tools to deal with each.

• Research Goals
  • Long-term Project Goals
    • Evaluate the viability of event time-difference and event pattern-occurrence as factors in behavior-based Intrusion Detection Systems for differentiating between human and automated program behavior.
    • In the future, determine how these factors can be added into Intrusion Detection Systems to help identify attackers swiftly.
  • Short-term Goals
    • Develop and finalize protocol for capture and analysis of honeypot machine-log data administered over by the National Center for Supercomputing Applications
      – Honeypots are a type of security architecture set up to gather information on malicious activity
    • Identify any trends or regular activity in the data

• Methods
  1. Organized honeypot log files by time/event/datatype
  2. Employed Syntactic Pattern Recognition of events in order to establish patterns
  3. Pulled CRON (known program) patterns/times/frequency to form control

• Preliminary Results
  • Protocol creation complete
  • Trends and Regular Activity
    • The entire network test group (n=63) averaged 4.67 ± 1.88.
    • The combined keystroke test group (n=190) averaged .26 ± .04 seconds.
    • The keystroke data revealed four unknown Pattern Groups, two of which were individual events.
    • While the network group had a total of seven unknown Pattern Groups, two of which were individual event occurrences.

• Conclusions
  • Some groups complete events within a rapid period of time, and repeat the same pattern of events over and over with little to no deviation.
  • Other groups take a longer period of time to complete events and fall outside the standard deviation.
  • This initial research has shown that Pattern_Occurrence and Time_Difference are indeed likely viable factors to separate human behavior from automated program behavior in an IDS and need further study.

• Future Research
  • Obtain larger sample size to replicate preliminary results and improve statistical significance
  • Establishing a way to add normalized human behavior data (as honeypots servers, by design, do not have regular users)
  • Designing an experiment to control for issues like distance-from-server lag, IP bounce, etc.

• Acknowledgments
  This project would not have been possible without help from my advisor, Dr. Masooda Bashir or assistance and data access from Alex Withers and the NCSA. Nor, without technical assistance from Bartosz G. Kosciarz and Seoung Kyun Kim. This material is based upon work supported by the Maryland Procurement Office under Contract No. H98230-14-C-0141.

• References
  [List of references]