Internet Ballistics: Retrieving Forensic Data From Network Scans

Bryan Parno, Harvard University
Tony Bartoletti, Computer Incident Advisory Capability (CIAC)
Lawrence Livermore National Laboratory

The typical network receives millions of hostile probes every day. A significant portion of these probes constitute network scans. During a network scan, the attacker sends connection requests to every possible network address and listens for replies indicating the presence of a (possibly vulnerable) computer. Since a network scan often serves as a precursor to an attack, reliable identification of scanners can significantly enhance cyber-security. Furthermore, the ability to map adversary hierarchies and correlate attacks with events in the real world contributes to counterintelligence work. For a variety of reasons, source IP addresses fail to provide the necessary identification information. However, analyzing packet arrival timing data reveals highly distinctive patterns that may correlate with the attacker's choice of tools, physical platform and/or network location. By selecting data transforms conducive to periodic analysis, we can use wavelet techniques to achieve over 1,000x compression ratio while still preserving the essential features. Initial experiments indicate our methods consistently identify patterns in the data. In future work, we plan to perform controlled scans using common network scanning tools from multiple locations to refine our identification techniques, allowing us to reliably fingerprint network scanners, without relying on the source IP address.