CS443 Paper Review
Lei Yang
2005-5-26

****

Title

Automated worm fingerprinting

Author

Sumeet Singh, Cristian Estan, George Varghese and Stefan Savage

Summary

Manual worm signature extraction is an expensive and slow procedure that can't possibly catch up with the speed worms propagate. This paper address the problem and describes Earlybird, a prototype system that automatically detects and contains new worms on the net using precise signatures.

Most important ideas

The main goal of this work, is to quickly detect new worms, thereby stop them from spreading and infect more hosts. Worms spread by transmitting packets through the net. More specifically, they tend to transmit same string within packets from many sources to many destinations. A basic assumption here is that all worms have some invariant content, thus invariant packets will appear frequently on the network. Moreover, the source and destination addresses for such invariant packets will roughly follow some sort of distribution. This work, therefore, identifies new worms by sifting through network traffic for content strings that are both frequently repeated and widely dispersed. More specifically, their prototype system monitors all network traffics, count the frequency of packet occurrences, and count the source and destination addresses.

As one might expect, monitoring, counting, and analyzing worm signature is very expensive: there are too much data to process. So the main problem in this work is to develop an efficient algorithm to analyze network traffic for prevalent and widely dispersed content strings. To estimate content prevalence, they examine all substrings of packets of a certain length, maintain circular buffer spanning all packets in a network flow, use polynomial-based fingerprint of strings, neutralize hash collisions with multi-stage filters, and include port and protocol for extra differentiation. To estimate address dispersion, they deploy multiple scaling bitmap counters. These methods are effective in decreasing memory usage and limiting the amount of processing. In the evaluation part, they showed that Earlybird is effective in detecting all known worms and some new worms. The CPU and memory requirements are relatively low.

Relevance

I'd say this is a very solid paper. The idea is good and their assumption is practical. In addition, the systems is proved to be working on real worms. They have, pointed out some limitations of the work to conclude the paper. First problem is, this solution would not work on worms with little or no invariant content. Second, smart worms may attempt to evade the monitoring through IDS evasion techniques. The idea, is definitely relevant today, it should have several potential extensions, for example, SPAM prevention, detection of massintrusion attempts, and etc.