

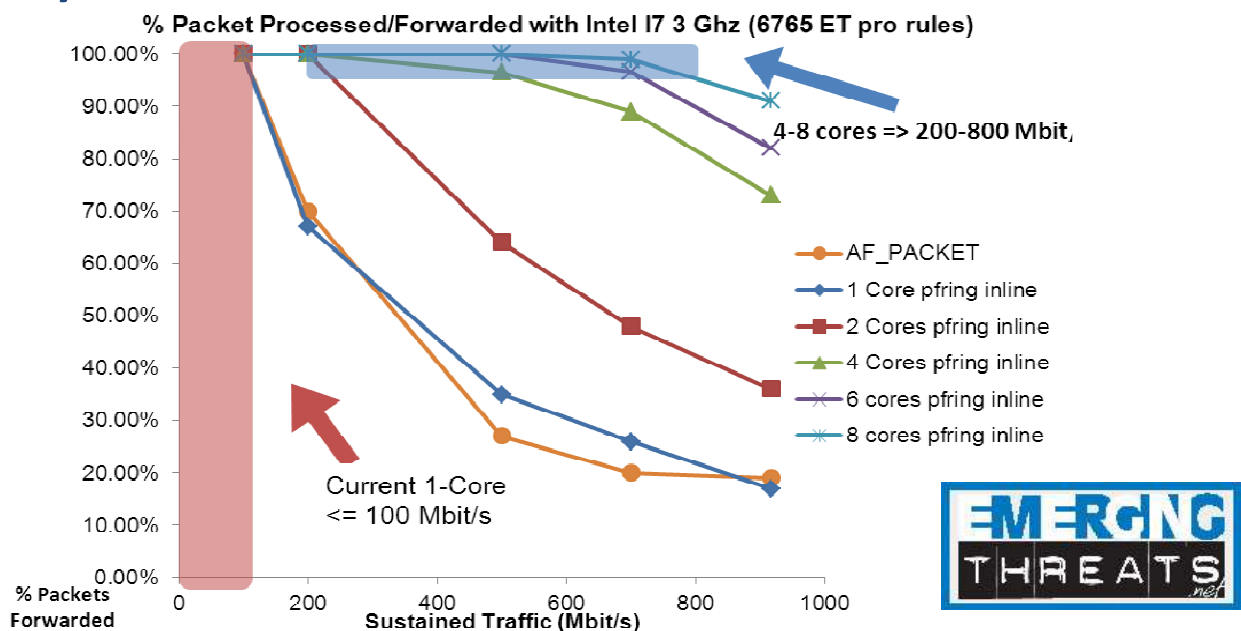
## How Metaflows Ranks Security Events

The Metaflows Security System (MSS) is software that monitors network traffic to detect and prevent cybersecurity incidents. The MSS also provides flow analysis, log management capabilities and reports on computer security policy violations.

One of the most common and significant problems in network security monitoring is false positive clutter. The MSS uses a complex and proprietary algorithm to prioritize true positives and deprioritize false positives.

The MSS uses a 3-layered system to address the problem of false positives.

### Layer 1: Session Level

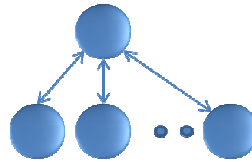


Traditionally, security events are generated by reconstructing a single session between two endpoints and finding known patterns that indicate a security violation within that session. This is the most basic level of intrusion detection and it is the extent of what most network security products offer today. It usually results in a very high false positive rate where important events are obfuscated by the huge volume of uninteresting security events. The MSS runs Snort as our session-level tool. Using a powerful open-source kernel module called PF-Ring, we have been able to parallelize Snort for both inline and passive applications. This allows performing session-level analysis on inexpensive off-the-shelf hardware. The MSS can process up to 800 Mbits of traffic with almost no packet drop using a standard, off-the-shelf machine costing approximately \$1000. The MSS can also process up to 5 Gbps using a standard, off-the-shelf machine with 2 Intel Xeon 6 core processors.

## Layer 2: Intra-Session

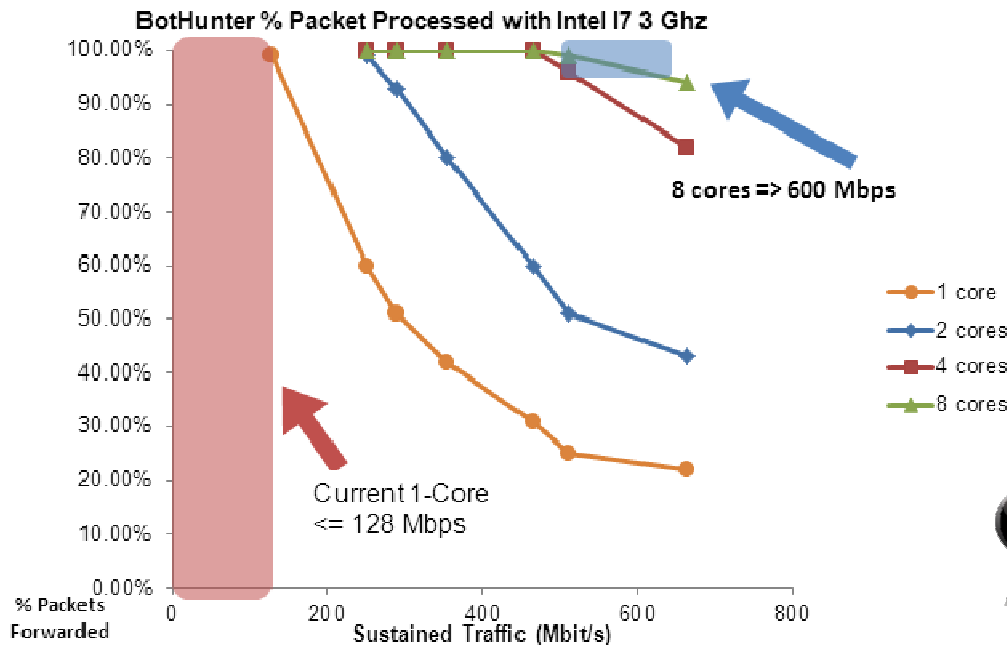
With intra-session correlation, we identify typical infection behavior by looking at alerts from multiple sessions belonging to a single home machine. The MSS positively scores alerts based on observing at least two events corresponding to the typical phases of a Bot Infection.

1. Inbound scanning
2. Exploit
3. Egg download
4. C&C communication
5. Outbound scanning/propagation

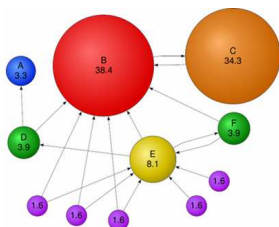


Intra-session analysis tends to eliminate false positives almost entirely and brings true positives to the forefront. This proprietary analysis is performed by Cyber-TA's BotHunter (licensed to Metaflows by SRI International). BotHunter intelligence feeds and rules are updated weekly from the SRI Malware Threat Center.

BotHunter also benefits from our multi-processing framework. The MSS can run BotHunter on as many cores as are available. As the number of cores increases, BotHunter's ability to process more network traffic increases dramatically.



## Layer 3: Intra-Domain (Predictive Global correlation)

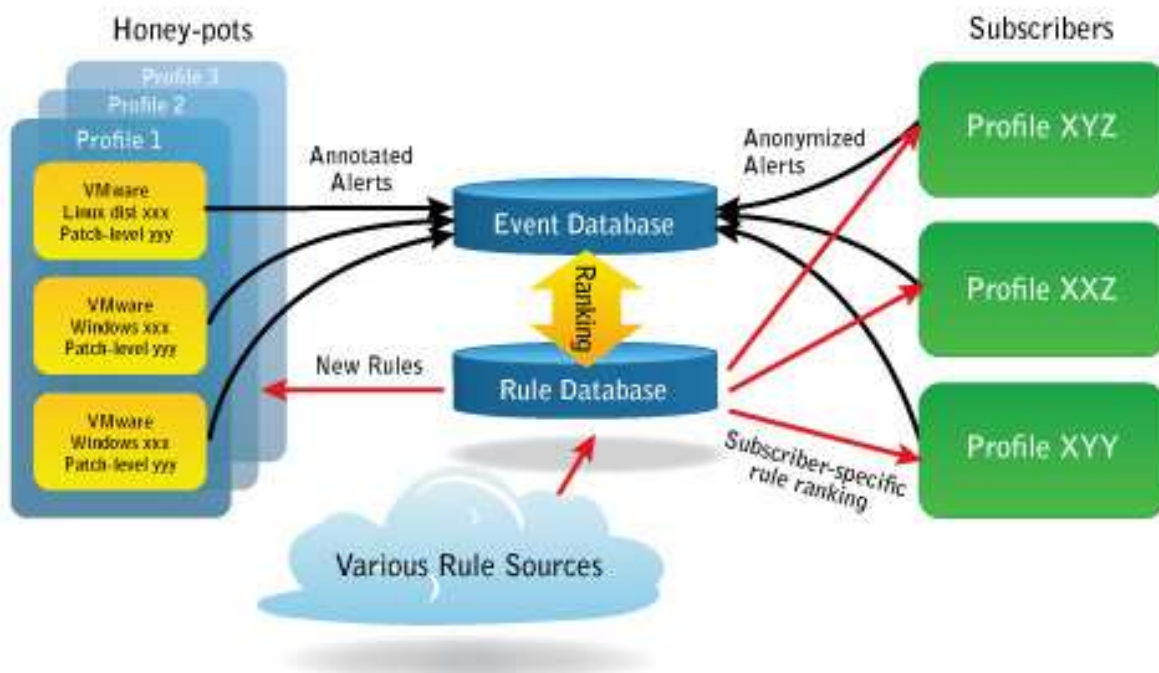


Research funded by the National Science Foundation has led to the developed of a proprietary intra-domain correlation algorithm that is mathematically similar to Google's Page ranking. Event scores are autonomously obtained from a global network of honeypot sensors monitored by the MSS. The honeypots are virtual machines that masquerade as victims. As the honeypots are repeatedly attacked, the MSS records both successful and unsuccessful hacker techniques and



corresponding security event information. The security events that trigger false positives are ranked negatively, thus providing insight into events that should be routinely ignored or turned off. Security events that trigger true positives are ranked positively thus improving their visibility. This information is then propagated in real time to each of our subscribers' sensors in the system to augment the session level and intra-session-level ranking described above. This additional intra-domain correlation is important because it adds operational awareness based on real-time intelligence.

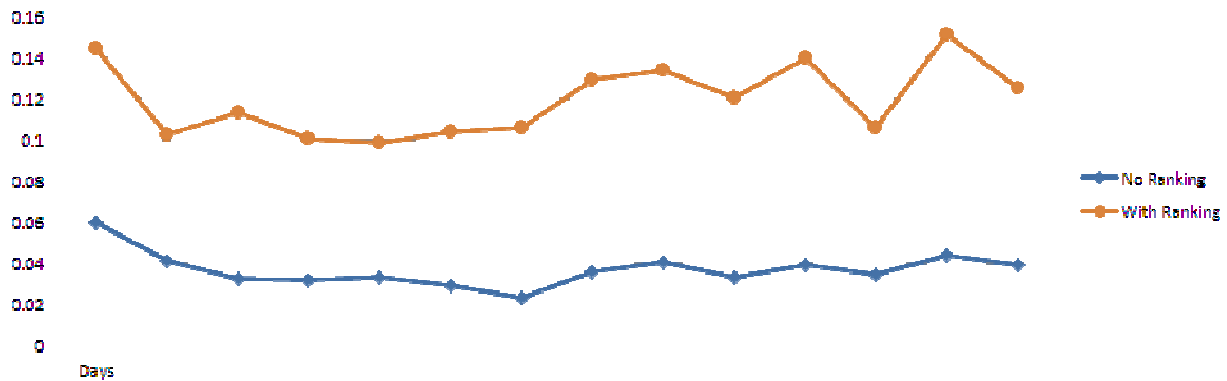
As shown in the figure below, honeypots work behind the scenes continuously mining global relevance data and flow intelligence (IP reputation) for threats that penetrate differing degrees of cyber defenses and different types of systems. Then, annotated data from all network sensors (whether they are honeypots or not) are compared and events are correlated with an algorithm similar to Google's page ranking ( $X = bs + aW*X$ ).



This process is designed to provide subscribers with intelligence data that takes into account the similarities and differences between the sources of that data. For space limitations we cannot explain the math and why it makes sense; however our system is very similar to the work described in [1] if you want to know more.

The outcome of the algorithm is that once a piece of intelligence reaches our system it is not equally distributed to all customers. Instead, it is mathematically weighted and routed to where it is most relevant, just as the first few web pages of a Google search yield the most relevant information for a particular search.

Our correlation engine also provides negative feedback to help de-prioritize and eventually mask/remove event signatures which cause false positives (negative ranking). In addition to real-time intelligence on true positive security events (positive ranking), our system also provides information on security alerts that are irrelevant by demoting them and reducing false positive clutter. In other words, this system can propagate known false positives and known true positives among sensors using a mathematical model that maximizes prediction.



The graph above quantifies the prediction power of the ranking algorithm. The experiment was carried out on the Snort event relevance data gathered between February 7 and February 22 2010. At the start of each day we performed the ranking operation over the previous day's Snort event data and compared the predicted ranking values with the actual events gathered that day from the sensors and honeypots. The simple prediction (blue line) is based on predicting that, for each sensor, the same event ranking is carried over from the previous day without running the algorithm (this is what people normally do today).

The Y axis (the hit ratio) is defined as the number of times the prediction matches the outcome in terms of the sign (positive or negative) divided by the number of non-zero ranking predicted.

- We increment the hit counter if the prediction and the outcome have both positive rankings.
- We increment the hit counter if the prediction and the outcome have both negative rankings.
- We decrement the hit counter if the prediction and the outcome have opposite signs.

The figure shows that the ranking prediction (orange line) is strictly superior to the simple day to day extrapolation by 141% to 350% (depending in the day). This might not seem too impressive on the surface but if you dig a little deeper this is what it means:



**MetaFlows.com / 877.664.7774**

**715 J ST STE 205 - San Diego - CA 92101**

- Assuming 5 minutes of human analysis time per incident, a system with no ranking would give you a hit rate that finds 1 actionable item every 20-30 incident investigations (or 0.4 incident per analyst hour).
- A system with predictive ranking would let you find 1 actionable item every 6-7 incidents investigations ( 2 incidents per analyst hour).

If you look at it that way, you can see why everyone using MetaFlows cannot stop using it.

[1] Highly Predictive Blacklisting, Jian Zhang, Phillip Porras, Johannes Ullrich SRI Interntional and the SANS Institute, Usenix Security, August 2008