Abstract. Commercially available Network Intrusion Detection Systems (NIDS) came onto the market over six years ago. These systems have gained acceptance as a viable means of monitoring the security of consumer networks, yet no commercial standards exist to help consumers understand the capacity characteristics of these devices. Existing NIDS tests are flawed. These tests resemble the same tests used with other networking equipment, such as switches and routers. However, switches and routers do not conduct the same level of deep packet inspection, nor require the higher-level protocol awareness that a NIDS demands. Therefore, the current testing does not allow consumers to infer any expected performance in their environment. Designing a new set of tests that is specific to the weak areas, or bottlenecks, of a NIDS is the key to discovering metrics meaningful to the consumers. Any consumer of NIDS technology can then examine the metrics used in the tests and profile his network traffic to these same metrics. The consumer can use standard test results to accurately predict performance on his network. This paper proposes a test methodology for standardized capacity benchmarking of NIDS. The test methodology starts with examining the bottlenecks in a NIDS, mapping these bottlenecks to metrics that can be tested, and then exploring some results from tests conducted.

1 Introduction and Scope

There are currently no industry standards for testing any aspect of Network Intrusion Detection Systems (NIDS). The NIDS industry is maturing along the same lines as the routers, switches, and firewalls that came before it, and has now reached the point where standardization of testing and benchmarking is possible. Attempting to define a testing standard is beyond the scope of this paper. Instead, the metrics and methodology used to properly verify the capacity of high speed NIDS are explored. Performance of NIDS is usually defined by false positive and false negative ratios, and speed or capacity. This paper addresses the issue of benchmarking the capacity of a NIDS. For the purposes of this paper we use capacity to refer to the ability of a NIDS to capture, process and perform at the same level of accuracy under a given network load as it does on a quiescent network.

Gauging the capacity of a NIDS is difficult. There are several variables in the characteristics of the network traffic that affect the performance of a NIDS. In the
last year there have been claims of NIDS performing at or near gigabit speeds. In every case, further investigation by reasonably sophisticated NIDS practitioners revealed critical flaws in the testing methodology.

The variety of technology used to perform network-based intrusion detection further complicates finding the proper metrics. The following technology is used for NIDS:
- Stateless inspection of the packets or packet headers
- Protocol decode and analysis
- Regular expression matching of packet data
- Anomaly detection

Most NIDS employ a mix of all of these methods. Some of the metrics discussed in this paper do not apply to all of the technologies. Choosing metrics and test methods valid for all NIDS in existence is impossible. Choosing a broad set of metrics that is generally applicable to most NIDS is possible. What are the proper metrics for performance testing? What testing methodology best evaluates these metrics? This paper focuses on these two questions.

The testing metrics and methodology described are intended for use on a NIDS located at the edge of an internal network functioning near the firewall or border router. The focus is further refined by looking at how these metrics apply to a NIDS using a combination of the technologies listed above. However, many of the metrics and methods included also apply to the performance of a NIDS inside the core of an enterprise network and to a NIDS employing other methods of detecting intrusions such as pure anomaly-based systems.

2 History

The majority of NIDS capacity benchmarks to date have been run by independent third parties either for publication in a trade magazine or at the request of the vendor for inclusion in marketing material. The test methodologies were developed based on experiences in the router- and switch-testing arenas.

These tests are generally not adequate for the purposes of developing a NIDS performance profile because the benchmark tests for switch and router capacity often forward packets of various sizes without regard for any protocol above IP or even the validity of the packets used. While routers and switches are typically not concerned with layer four and above, NIDS may discard packets that are not interesting. A NIDS also needs to look much deeper in a packet than a switch or a router to follow layer four and above. For example, a NIDS may discard TCP streams that are not opened using a valid three-way handshake. If a switch or router test is used the majority of the traffic might be ignored. The NIDS then performs very little deep packet inspection.

Since the results of a NIDS performance test based on these types of test methodologies are often skewed in the favor of the vendor, a consumer may believe these results are valid for his deployment and encounter strikingly different performance characteristics once the NIDS is fielded on his network.