Deciphering Detection Techniques: Part II Anomaly-Based Intrusion Detection

By Dr. Fengmin Gong, Chief Scientist, McAfee Network Security Technologies Group

March 2003
## Table of Contents

I. Introduction ............................................................................................................................................................................................................................ 3

II. What Is Anomaly-Based Intrusion Detection? .................................................................................................................................................................................................................................................. 3

   Anomaly-Based vs. Signature-Based  What’s the Difference? ................................................................................................................................................................................................................................................. 3

   It’s All in the Profile ................................................................................................................................................................................................................................................................................................................ 4

III. When Is Anomaly-Based Detection Most Useful? ...................................................................................................................................................................................................................................................... 4

   The Drawbacks of Using Anomaly-Based Detection Alone ........................................................................................................................................................................................................................................................................................................ 5

IV. What Are the Applicable Areas of Protection? ........................................................................................................................................................................................................................................................................ 5

   1. Misuse of Protocol and Service Ports ........................................................................................................................................................................................................................................................................ 5

   2. DoS Based on Crafted Payloads ........................................................................................................................................................................................................................................................................... 5

   3. DoS Based on Volume (DDoS/DRDoS) .......................................................................................................................................................................................................................................................................... 5

   4. Buffer Overflow ................................................................................................................................................................................................................................................................................................. 6

   5. Other Natural Network Failures ............................................................................................................................................................................................................................................................................ 6

V. Anomaly Detection Techniques ........................................................................................................................................................................................................................................................................... 6

   Protocol Anomaly Detection .................................................................................................................................................................................................................................................................................. 6

   Application Payload Anomaly ............................................................................................................................................................................................................................................................................. 7

   Statistical Anomaly—Statistical DDoS ........................................................................................................................................................................................................................................................................... 7

VI. The McAfee IntruShield Approach to Intrusion Detection ..................................................................................................................................................................................................................................................................... 8

VII. About the Author ........................................................................................................................................................................................................................................................................ 9

VIII. About McAfee Network Protection Solutions ...................................................................................................................................................................................................................................................................... 9

   McAfee IntruShield ................................................................................................................................................................................................................................................................................... 9

IX. About Network Associates ........................................................................................................................................................................................................................................................................ 10
I. Introduction

Hackers are smart, creative people. They have spawned hundreds of attacks designed to exploit known vulnerabilities in your machines and applications. Many of these exploits are so common that they are often referred to as "well-known attacks." For these attacks, your intrusion detection system undoubtedly uses a database of defined signatures for matching strings against these well-known attacks. But the clever hacker knows this, and continues to beat your IDS by crafting attack variants to beat your signature strings, or developing attacks that exploit new vulnerabilities. Where signature-based detection comes up short, anomaly-based detection closes the holes.

Anomaly-based intrusion detection is a newer method in the fight against exploits and misuse. By itself, anomaly-based detection is not a cure-all. But when used in conjunction with an effective signature-based detection solution, anomaly-based detection is a viable and effective means of protecting your network infrastructure and your company’s ability to do business.

This white paper details:

- An explanation of anomaly-based intrusion detection
- The situations in which anomaly-based Intrusion detection work best to protect your network
- The McAfee IntruShield System approach to anomaly-based intrusion detection

II. What Is Anomaly-Based Intrusion Detection?

Simply put, anomaly-based intrusion detection triggers an alarm on the IDS when some type of unusual behavior occurs on your network. This would include any event, state, content, or behavior that is considered to be abnormal by a pre-defined standard. Anything that deviates from this baseline of "normal" behavior will be flagged and logged as anomalous. "Normal" behavior can be programmed into the system based on offline learning and research or the system can learn the "normal" behavior online while processing the network traffic.

Some examples of anomalous behavior include:

- HTTP traffic on a non-standard port, say port 53 (protocol anomaly)
- Backdoor service on well-known standard port, e.g., peer-to-peer file sharing using Gnutella on port 80 (protocol anomaly and statistical anomaly)
- A segment of binary code in a user password (application anomaly)
- Too much UDP compared to TCP traffic (statistical anomaly)
- A greater number of bytes coming from an HTTP browser than are going to it (application and statistical anomaly)

Anomaly-Based vs. Signature-Based  What’s the Difference?

When a network is being monitored for potential security incidents, an IDS can implement anomaly and/or signature-based intrusion detection. There are advantages and disadvantages to each method—the best-fortified network uses the two methods together to provide the maximum defense for the network infrastructure.

A signature generally refers to a set of conditions that characterizes the direct manifestation of intrusion activities in terms of packet headers and payload content. Historically, the signature-based method has been the more common of the two methods when looking for suspicious or malicious activity on the network. This method relies on its database of attack signatures and when one or more of these signatures match what’s observed in the live traffic, in the case of a NIDS, an alarm is triggered and the event is logged for further investigation. Signature-based intrusion detection is only as good as its database  if a signature is not in the database, the IDS will not catch the attack. This is obviously a drawback when you consider that hackers spend a great deal of their time crafting attacks designed to fool signature-based systems.
Anomaly-based intrusion detection, on the other hand, takes a more generalized approach when looking for and detecting threats to your network. A baseline of “normal” behavior is developed, and when an event falls outside that norm, it is flagged and logged. The behavior is a characterization of the state of the protected system, which is both reflective of the system health and sensitive to attacks. In this context, an anomaly-based method of intrusion detection has the potential to detect new or unknown attacks. Like the signature-based method, however, anomaly-based intrusion detection also relies on information that tells it what is normal and what isn’t. This is called a profile, and it is key to an effective anomaly-based intrusion detection system.

It’s All in the Profile

For anomaly-based intrusion detection to be effective, it must have a robust profile that characterizes normal behavior. The target could be a host/IP address, VLAN or physical LAN segment. A profile consists of a comprehensive list of parameters and values that are specifically geared to the target being monitored. A robust profile must be stable and consistent in tracking the normal behavior of the target environment. An effective anomaly profile must also be sensitive to occurrences of any events that are deemed to be security concerns. Constructing an effective profile involves gathering information on behavior and activity currently considered acceptable on your network. Profiles can vary in complexity from a couple of simple thresholds to comprehensive content characterizations to multi-variable distributions. For example, a profile could include following types of information:

- Occurrence patterns of specific commands in application sessions, e.g., a successful remote login session should contain authentication commands
- Association of content types with different fields of application protocols, e.g., by design the password for an application must be 7-bit ASCII of 8 to 64 characters
- Connectivity patterns between protected servers and the outside world (protocol and port breakdown, volume, and duration), e.g., by design the only expected network activities for a dedicated Oracle database server are transactions with three known Oracle thin clients
- Rate and burst length distributions for all types of traffic, e.g., on a normal internal segment, you do not see fragmented IP packets bursting more than 1 minute at 100 packets per second

In addition to being robust and sensitive, profiles based on network traffic should also be adaptive and self-learning. Adaptive profiles can account for normal network changes to avoid raising false alarms. Self-learning is critical to ensure wide and successful deployment of anomaly-based detection mechanisms. First of all, it is generally very difficult for security analysts to manually set the profiles because of the complexity and dynamically changing traffic statistics. Furthermore, while it is highly desirable to apply anomaly-based detection at various levels of traffic aggregation to achieve the most accurate protection, (e.g., applying it to the traffic of the whole finance department, that of the server farm and that of a single critical server) accurate traffic statistics at all these levels are nonexistent in most of the organizations. With self-learning, the detection mechanism can learn the normal behavior of the traffic stream assigned to it and provide detection based on the learned profile.

III. When Is Anomaly-Based Detection Most Useful?

Because it detects any traffic behavior that is new or unusual, the anomaly-based method is excellent at providing early warning for potential intrusions. These warnings can cover stealthy reconnaissance attempts, backdoor activities, and certain natural failures in the network. Additionally, anomaly-based intrusion detection is useful for detecting these types of attacks:

- New buffer overflow attacks carrying shellcode
- New exploits
- Intentionally stealthy attacks (e.g., using ADMutate to transform a shellcode)
- Variants of existing attacks in new environments (e.g., worms using different file names as they propagate)
Keep in mind that anomaly-based intrusion detection provides the best protection for your network when it is combined with a signature-based approach. For known attacks, signatures generally provide the most accurate detection in shortest time. In this way, known attacks as well as those attacks whose signatures are not yet known and attacks that exhibit modified behavior can be blocked.

**The Drawbacks of Using Anomaly-Based Detection Alone**

Not every anomaly is an intrusion or policy violation. For the security professional, the detection results of an anomaly-based IDS will require additional resources, due to the following issues:

- Anomaly-based IDS are more prone to generating false positives due to the ever-changing nature of networks, applications and exploits.
- Due to aggregation and abstraction used in profiling, alerts generated may not contain detailed enough information for forensic analysis and hinder the development of counter-measures.
- A longer detection and alerting time due to a minimum observation window for determining anomalies to be intrusions. For example, an exploit using a missing command in a session can only be identified when a session is completed and will necessitate keeping track of state and context longer than for a signature matching. Often, this could be too late.

**IV. What Are the Applicable Areas of Protection?**

When examining information gathered by anomaly-based intrusion detection, the focus shifts to the measurable effects of the event, rather than the mechanics of the event’s execution. This is applicable in those areas of protection where measurable effects are most meaningful and where anomaly-based intrusion detection is most helpful.

1. **Misuse of Protocol and Service Ports**

For a given standard protocol, someone invokes either experimental or obsolete features of the protocol. Sometimes, this is an indicator of malicious activity, in other cases, this may be someone inside your network performing testing or research. A protocol can also be misused when an attacker modifies it in order to tunnel through a firewall. Installation of backdoor services on well-known standard ports is another common misuse of service ports.

2. **DoS Based on Crafted Payloads**

When a malicious intruder creates an attack using a crafted IP packet, the resulting Denial of Service (DoS) can occur on the network bandwidth, CPU cycles, memory resources, or application process/programs. Examples of this type of DoS include process table exhaustion, IP stack crashing, or a Web application “soft spot.” The observable impact of this DoS attack will be an anomaly in service quality.

3. **DoS Based on Volume (DDoS/DRDoS)**

When attacks have progressed beyond control channel activity, anomaly-based intrusion detection is the only reliable means for detection in the case of the DoS attack that floods the network with a large volume of traffic. This is because sophisticated attack traffic may not be distinguishable from regular network traffic on an individual packet basis and the attack does not manifest a specific signature that can be captured by signature-based mechanisms.

For example, the following traffic pattern anomalies can be observed as a result of the Distributed Denial of Service (DDoS):

- TCP control packet statistics for TCP SYN flood
- Relative volumes of TCP, UDP, and ICMP traffic for UDP or ICMP flood
• Statistics on HTTP request vs. response balance for DDoS on the server via normal-looking requests; a difference in request vs. response packet counts, a difference of request vs. response byte counts, and the time interval between request and response.

The anomalies that are seen as a result of the DoS attack based on volume are applicable for attacks from outside to inside (O2I), from inside to outside (I2O), from inside to inside (I2I).

4. Buffer Overflow

The buffer overflow is the most prevalent vulnerability exploited by attackers. Buffer overflow with shellcode execution is the most serious form of this exploit because a successful attack can result in arbitrary program execution on the victim system(s).

Many exploited fields, such as user passwords for FTP, are supposedly made of printable ASCII characters based on the standard—Request For Comments (RFCs) by the Internet Engineering Task Force (IETF).

Excessive non-printable ASCII characters are anomalies of strong suspicion. Furthermore, shellcode embedded in these fields are sure signs of malicious intent.

5. Other Natural Network Failures

Subtle failures in routers/switches can result in changes in traffic pattern observed at certain points of the network. This can be in the form of sudden drop in the volume of traffic due to broken connections, or in the form of traffic shift from one link to another due to traffic rerouting as a recovery action. All these changes are noteworthy and can be detected as traffic anomalies.

V. Anomaly Detection Techniques

Anomaly-based detection represents a broad spectrum of detection techniques. While all these techniques use certain levels of abstraction in order to provide a broad coverage of security events, they may differ in many ways. There are different types of measures to use depending on what is being protected and what is being monitored. One can define profiles in terms of simple thresholds or more complex statistical distributions; and profiles can be self-learned or manually set, adaptive, or static. In the absence of a universally agreed taxonomy, three broad types are discussed in the following paragraphs.

Protocol Anomaly Detection

In this white paper, protocol anomaly refers to all exceptions related to protocol format and protocol behavior with respect to common practice on the Internet and standard specifications. This includes network and transport layer protocol anomalies in layers 3-4 and application layer protocol anomalies in layers 6-7. There are many aspects of the basic TCP/IP stack behavior that should be monitored. This can be done by performing full IP defragmentation, TCP reassembly, and checking for any unusual conditions in the process. Furthermore, many exceptions that could lead to ambiguous interpretation by the end hosts can be safely removed when the IDS system is inline. This has been referred to in the IDS world as traffic normalization or protocol scrubbing.

Some examples of anomalies detectable through layer 3-4 protocol anomaly detection include the following:

• IP fragmentation overlaps and suspicious IP options
• Unusual TCP segmentation overlaps and illegal TCP options and usage
• Corrupt checksums
• Length and flag consistency
When an IDS is monitoring application protocol behavior, it must be able to perform deep application protocol parsing, which is also known as decoding. The following anomalies are examples of protocol anomalies that could be detected when application protocol behavior is being observed:

- Illegal field values and combinations
- Illegal command usage
- Unusually long or short field lengths, which can indicate an attacker is attempting to introduce a buffer overflow
- Unusual number of occurrences of particular fields/commands
- Running a protocol or service for a non-standard purpose or on a non-standard port

**Application Payload Anomaly**

Application anomaly must be supported by detailed analysis of application protocols. Only with deep analysis will it be possible to separate different logical fields and define accurate behavior constraints for them. Application anomaly also requires understanding of the application semantics in order to be effective. For example, one needs to know what type of encoding is legal for a given field, and what other applications can be embedded within it. One good example of application level anomaly is the presence of shellcode in unexpected fields. A reliable anomaly profile allows shellcode execution attacks to be detected without knowing what particular exploit code is involved, or even the existence of exploit code.

**Statistical Anomaly—Statistical DDoS**

To fully characterize the traffic behavior in any network, various statistical measures are used to capture this behavior. For example, normal TCP traffic follows a well-defined three-way handshake process for connection setup, goes through data transfer phase, and then completes the communication by tearing down the connection. At any given point of observation on a network, there is a stable balance among different types of TCP packets in the absence of attacks. This balance can be learned and compared against short-term observations that will be affected by attack events. Additionally, the statistical algorithm must recognize the difference between the long-term (assumed normal) and the short-term observations in a given protected environment to avoid generating false alarms on normal traffic variations. Another type of measure could capture the intensity of the traffic monitored. While time-of-the-day or day-of-the-week variations in traffic volume are normal and well understood, traffic rate distributions on a multi-week scale are fairly stable for any normal network environments.

Profiles based on statistical measures could raise DDoS anomalies based on rare events of the difference between the long- and short-term distributions or based on a rare occurrence of long bursts of high-rate traffic. A well-designed system would allow the user to set a sensitivity level to reflect how tolerant their network or servers are to traffic surge. The lower the sensitivity level, the more severe the traffic profile deviation must be before the algorithm raises a DDoS alarm. The normal profiles are continuously learned while the system is in detection mode, with safeguard against statistics poisoning under attacks. This allows the anomaly profiles to adapt to typical environmental changes that occur in an organization.

Combining statistical measures can allow further differentiation among different traffic-impacting events, such as a real volume-based DDoS, versus a harmless flash-crowd condition, or sudden route change. These conditions do occur in real-life networks and it is important that users are provided as specific information as possible when alarms are raised. In order to make sure that DDoS attacks are detected regardless of what type of packets are used, anomaly profiles must be comprehensive enough to account for every possible packet on the network. For example, some of the events that can be detected by the IntruShield DDoS profiles include:

- SYN Flood attacks
- UDP Flood attacks
- ICMP Flood attacks
- TCP data segment flood attacks
VI. The McAfee IntruShield Approach to Intrusion Detection

The industry’s first real-time network intrusion detection and prevention platform has taken intrusion detection to a new level. McAfee IntruShield network security products deliver an integrated hardware and software solution, which delivers comprehensive detection and protection from known, first strike (unknown), DoS, and DDoS attacks from several hundred Mbps to multi-gigabit speeds.

The award-winning IntruShield architecture integrates patented signature, anomaly, and Denial of Service detection on a single purpose-built appliance. This not only enables highly accurate detection, but also empowers administrators with smart tools and processes, and enables flexible and scalable deployment for global businesses and vital government agencies.

The IntruShield architecture employs a combination of threshold-based and patented self-learning, profile-based detection techniques that delivers unmatched intelligence to intrusion detection.

With straightforward threshold-based detection, IntruShield administrators can configure data traffic limits to ensure their servers will not become unavailable due to overload. Meanwhile, self-learning methodologies enable IntruShield to study the patterns of network usage and traffic over time; thus understanding the wide variety of lawful, though unusual, usage patterns that may occur during legitimate network operations. Highly accurate detection techniques are essential because popular Web sites and networks do experience legitimate and sometimes unexpected traffic surges during external events. Critical factors that make effective anomaly-based protection possible with IntruShield are:

1. **Recognize anomaly does not equal attack:** The IntruShield system detects all anomalies and differentiates between anomalies that are real attacks so that users can make informed decisions.

2. **Adaptive, Granular and Robust Profile:** Highly accurate and self-learning profile that can be applied at any level of traffic granularity to detect anomalies. The target profile can be intelligently learned for a single IP address or CIDR block, VLAN, or LAN segment within hours. A targeted and robust profile in conjunction with broad anomaly coverage can reduce false positives and false negatives. For example, one can choose to apply detection profiles to a single server in a large server farm or a physical LAN segment. This allows the IntruShield IDS to detect and potentially block a low-bandwidth DoS attack that would be missed by a legacy anomaly-based IDS that looks for anomalies at an aggregate bandwidth level.

3. **Comprehensive anomaly detection:** IntruShield employs statistical, protocol, and application anomaly detection, with a comprehensive set of measures to account for all traffic types on the network. This allows the IntruShield IDS to detect the following types of intrusions:
   - New Buffer-overflow attacks carrying shellcode
   - Use of modified protocols for tunneling through firewalls (e.g., Gnutella on port 80)
   - Backdoor intrusions
   - DoS/DDoS attacks
   - Low-bandwidth DoS/DDOS attacks
   - Intentionally stealthy attacks
   - Variants of existing attacks in new environments
   - Incoming versus Outgoing attacks

4. **Recognize the complementary roles between anomaly-based and signature-based detection methods:** The IntruShield IDS performs deep-packet inspection and full protocol decode/analysis for a large number of network and application protocols, thereby providing highly accurate context and statistics to the anomaly-based detection engine. The anomaly-based detection engine is integrated into the IntruShield architecture, and cross correlation between anomaly events and the stateful signature checks allow the best coverage and accuracy for intrusion detection.
5. **Purpose-built appliance power:** IntruShield monitors events across all protocol layers, with full state, and without ambiguity at up to 2Gbps; it discovers and analyzes protocols regardless of whether or not standard ports are used. Without the purpose-built platform, it will not be possible for the anomaly algorithms to see everything completely and accurately for accurate protection.

**VII. About the Author**

Dr. Fengmin Gong is the Chief Scientist for the McAfee Network Security Technologies Group, where he is responsible for driving the continued innovation of IntruShield’s security architecture - leveraging his expertise in areas such as signature, anomaly, and denial of service detection. Before to his work on IntruShield, Dr. Gong was Director of Advanced Networking Research at MCNC, a provider of sophisticated electronic and information technologies and services aimed at businesses and government agencies.

While at MCNC and earlier at Washington University, he was involved in advanced security and networking projects for agencies such as DARPA, NSA, NSF, NLM, and NASA. During his time at MCNC he was also Adjunct Assistant Professor of Computer Science at North Carolina State University.

In a distinguished academic and research career, Dr. Gong has written and contributed to nearly forty research papers on network intrusion, anomaly detection, secure collaboration, multi-media content delivery, and network quality of service. Dr. Gong has presented his research at industry events such as IEEE technical forums, as well as SIGGRAPH, DISCEX, NOMS, and ISCEX.

**VIII. About McAfee Network Protection Solutions**

McAfee Network Protection Solutions keep both large and smaller distributed networks up and protected from attacks. Best-of-breed network protection solutions in the portfolio include the Sniffer Network Protection Platform for performance management and fault identification, InfiniStream performing security forensics on network activity, Network Performance Orchestrator (nPO) for centralizing and managing network activity, and McAfee IntruShield delivering network-based intrusion prevention.

**McAfee IntruShield**

McAfee IntruShield, a part of Network Associates’ McAfee Network Protection Solutions family of products, is a unique cutting-edge technology that prevents intrusions "on the wire" before they hit critical systems. Highly automated and easily managed, McAfee IntruShield is designed with such flexibility that it can be implemented in a phased approach - that overcomes the false positives inherent with today’s legacy intrusion detection systems - and thus enables you to develop the right policy for blocking in your unique IT infrastructure. For example, you can deploy in-line to notify and block known attacks, and to notify-only on unknown attacks. Or you can implement complete blocking but just for business-critical network segments. IntruShield is delivered in a high-speed appliance which is able to scan traffic and assess threat levels with blinding speed, even on gigabit networks. It can be used at the edge or in front of key "core" resources. IntruShield has been crafted to satisfy both the security and network administrators as it stops a wide range of network attacks but does so with network latencies typically less than 10 milliseconds. IntruShield also looks for anomalous behavior and includes specialized analysis to find new denial of service "mass attacks".
IX. About Network Associates

With headquarters in Santa Clara, Calif., Network Associates, Inc (NYSE: NET) creates best-of-breed computer security solutions that prevent intrusions on networks and protect computer systems from the next generation of blended attacks and threats. Offering two families of products, McAfee System Protection Solutions, securing desktops and servers, and McAfee Network Protection Solutions, ensuring the protection and performance of the corporate network, Network Associates offers computer security to large enterprises, governments, small and medium sized businesses, and consumers. These two product portfolios incorporate Network Associates’ leading McAfee, Sniffer and Magic product lines. For more information, Network Associates can be reached at 972-963-8000 or on the Internet at http://www.networkassociates.com/.

All Network Associates® products are backed by our PrimeSupport® program and Network Associates Laboratories. Tailored to fit your company’s needs, PrimeSupport service offers essential product knowledge and rapid, reliable technical solutions to keep you up and running. Network Associates Laboratories, a world leader in information systems and security, is your guarantee of the ongoing development and refinement of all our technologies.

Network Associates, Sniffer, McAfee, Magic Solutions, InstruShield, and PrimeSupport are either registered trademarks or trademarks of Network Associates, Inc. and/or its affiliates in the US and/or other countries. Sniffer® brand products are made only by Network Associates, Inc. All other registered and unregistered trademarks in this document are the sole property of their respective owners. ©2003 Networks Associates Technology, Inc. All Rights Reserved.