A Publicly Available Data Set for the Evaluation of Signature-Based IDS

Frédéric Massicotte networksystems-security@crc.ca
Network Security Group
Communication Research Center Canada

François Gagnon fgagnon@sce.carleton.ca
Network Management & Artificial Intelligence Laboratory
Carleton University Canada

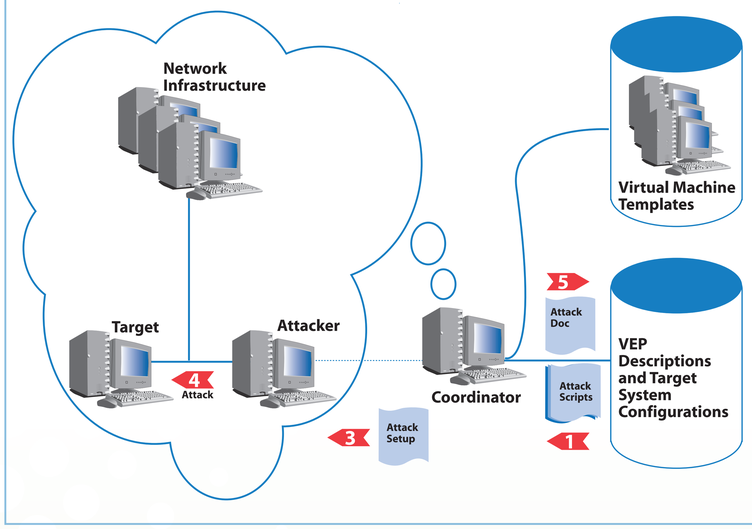
Motivations

IDS evaluation data sets have many problems: availability, age, documentation, updatability, …

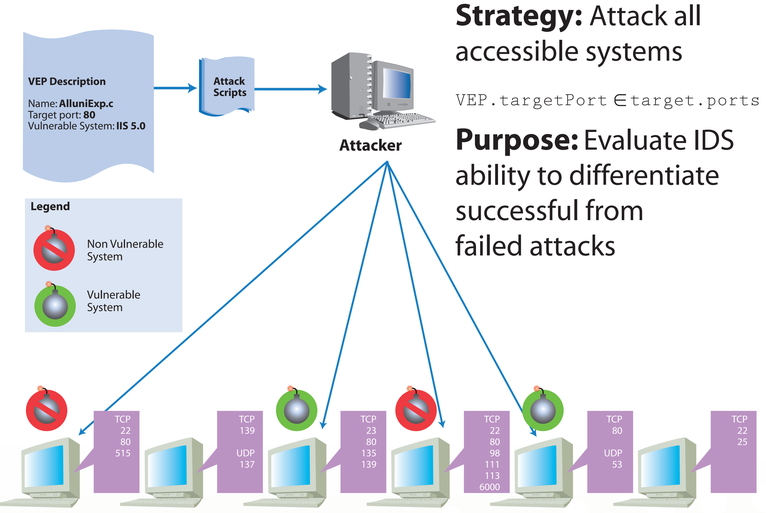
- Many data sets are proprietary.
- Others are old and obsolete, but still used today.
- Data sets are not properly documented.
- Most data sets cannot be easily updated.

We created a publicly available data set.

Automatic Generation of the Data Set



Data Set Generation



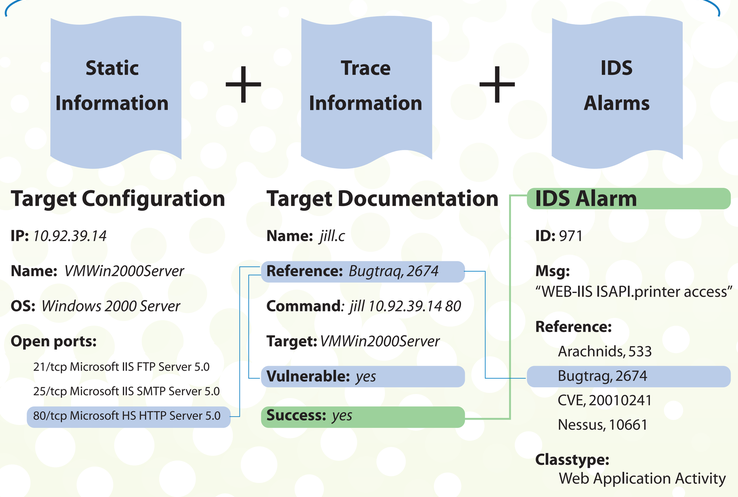
Strategy: Attack all accessible systems

`VEP.targetPort ∈ target.ports`

Purpose: Evaluate IDS ability to differentiate successful from failed attacks

Data Set Documentation

Automatic IDS Evaluation?



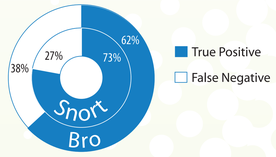
Data Set

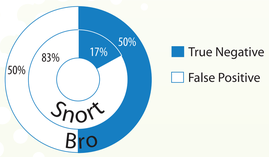
- Composed of more than 7 000 documented attack scenarios each containing:
  - -The attack traffic trace in pcap format
  - -The attack execution parameters in xml format
  - -The attack command ouput in text format
- 94 Vulnerability Exploitation Programs (VEP)
  - -covering 49 different vulnerabilities distributed over 17 ports
- 108 different target systems
  - -Windows 95, 98, 2000, XP, 2003
  - -Linux RedHat and Linux Suse
  - -FreeBSD, NetBSD and OpenBSD

IDS Evaluation with the Data Set

Tested IDS:

- Snort 2.3.2
- Bro 0.9a9 with Snort 2.3.2 rules converted using s2b





Successful Attacks

- Snort detects more attacks
  - -Bro is not able to translate all Snort plugins
- Both IDS miss more than 25% of the attacks

Failed Attacks

- Bro is better to detect failed attacks
  - -Brolooks at error messages and server identity
- Both IDS are still very noisy

Communications Research Centre Canada
An Agency of Industry Canada
Centre de recherches sur les communications Canada
Un organisme d'Industrie Canada
Canada