GEORGIA TECH INFORMATION SECURITY CENTER

# BotHunter

## Detecting Malware Infection Through IDS-Driven Dialog Correlation


**Guofei Gu**[1], Phillip Porras[2], Vinod Yegneswaran[2], Martin Fong[2], Wenke Lee[1]

[1]College of Computing
Georgia Institute of Technology

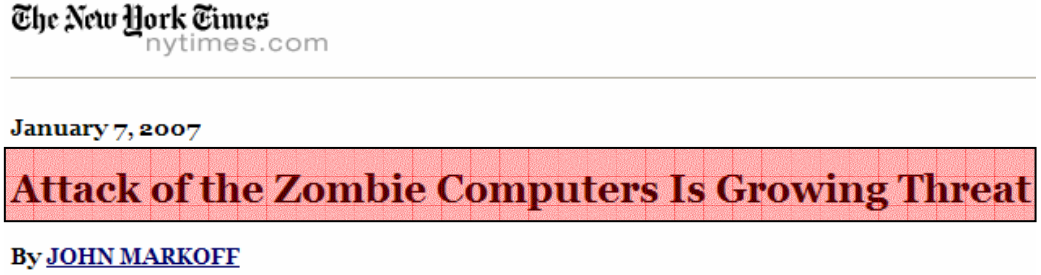
[2]Computer Science Laboratory
SRI International


Georgia Tech College of Computing

SRI International

# Botnets Are Emerging Threats to Internet Security



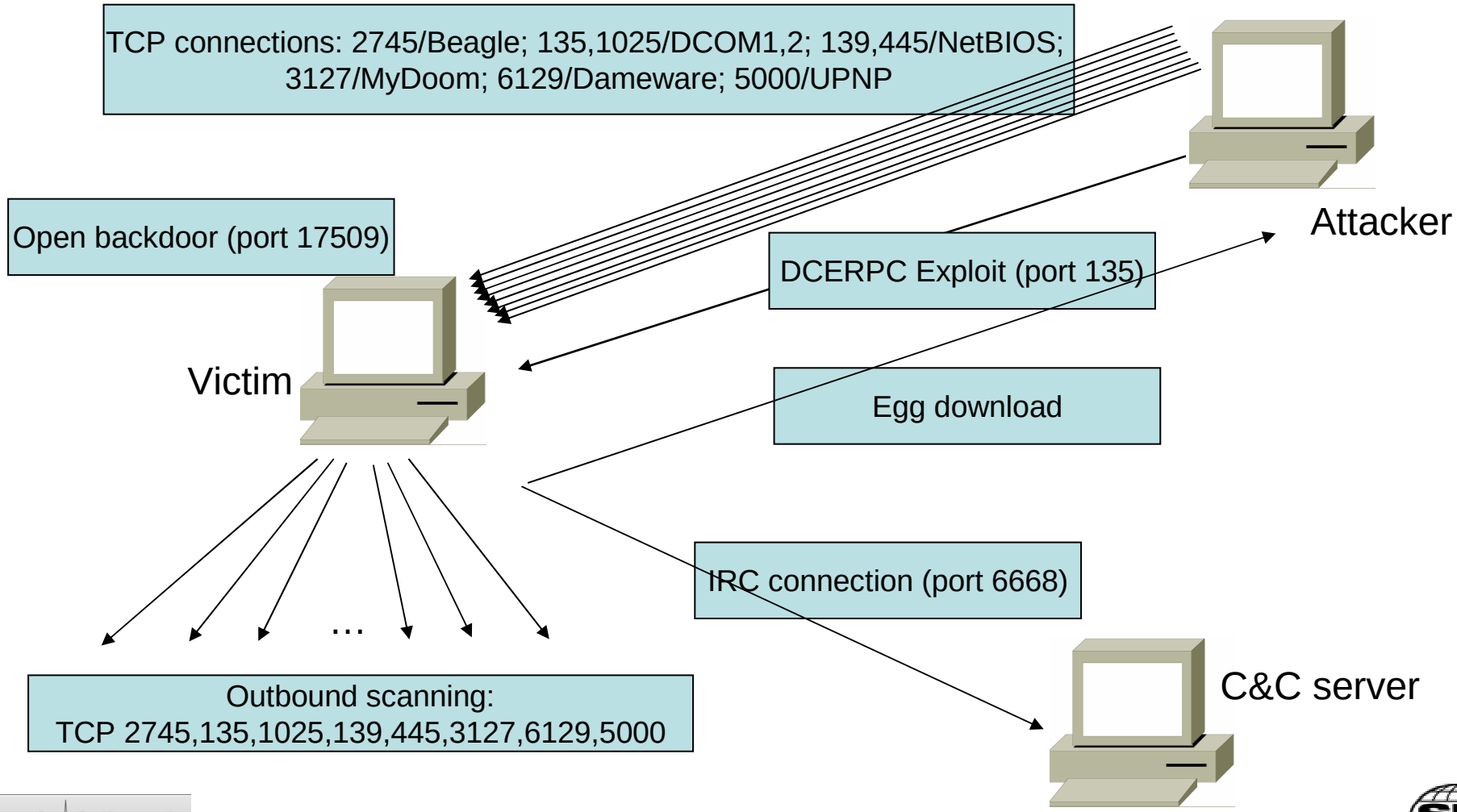
# What Are Bots & Botnets Today?

- Bots: malware that has
  - a remote control facility (C&C)
    - IRC, HTTP, P2P
  - a spreading mechanism to propagate
    - Remote vulnerability scan, Email, Drive-by download, IM
- Botnets - networks of bots
- Bots/Botnets are used for
  - DDoS, Spam, Click fraud, Data theft, …

# Detecting Bots/Botnets Is Difficult

- Bots are actively evolving
  - Infection vectors
  - Binary updates
  - C&C servers/communications
  - Scanning strategies
- Traditional IDSs/IPSs are less helpful in identifying bots (too many false positives)
- Only looking at one specific aspect is probably not enough

# Bot infection case study: Phatbot

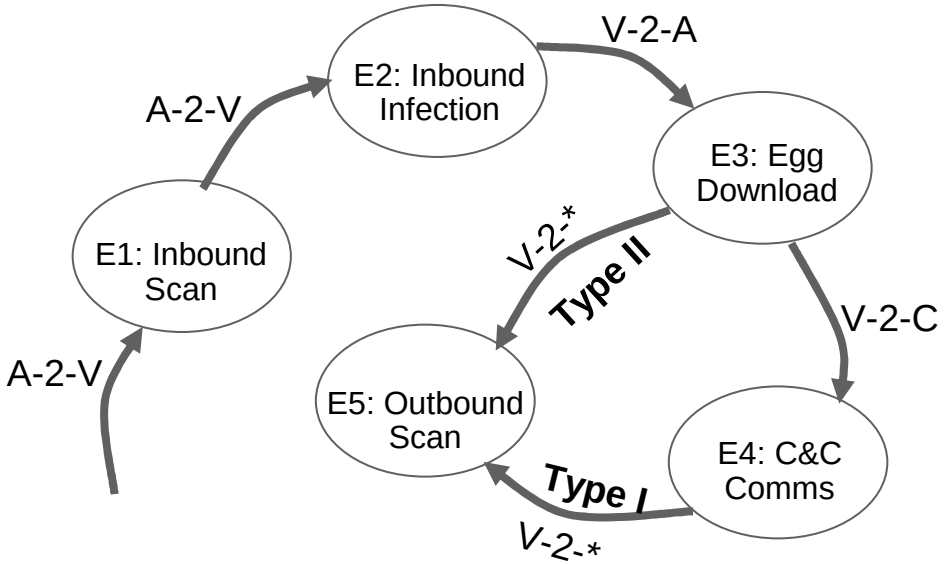


# What Is BotHunter?

- **BotHunter** – an IDS-Driven ***Dialog*** Correlation Engine
  - Protect enterprise/campus network
  - Monitors **<u>two-way communication</u>** flows between internal networks and the Internet for signs of bot and other malware
  - Correlates dialog trail of inbound intrusion alarms with outbound communication patterns
  - Produces a comprehensive 'bot' *Profile* that captures
    - Infection source/methods/pattern
    - identity of the locally infected host
    - most likely C&C address
    - all related dialog warning summaries

# Dialog-based Correlation

BotHunter employs an
***Infection Lifecycle Model***
to detect host infection behavior




- Egress point (internal – external)
- Search for duplex communication sequences that map to I.L. model
- Stimulus does not require strict ordering, but does require temporal locality

# BotHunter - Correlation Framework

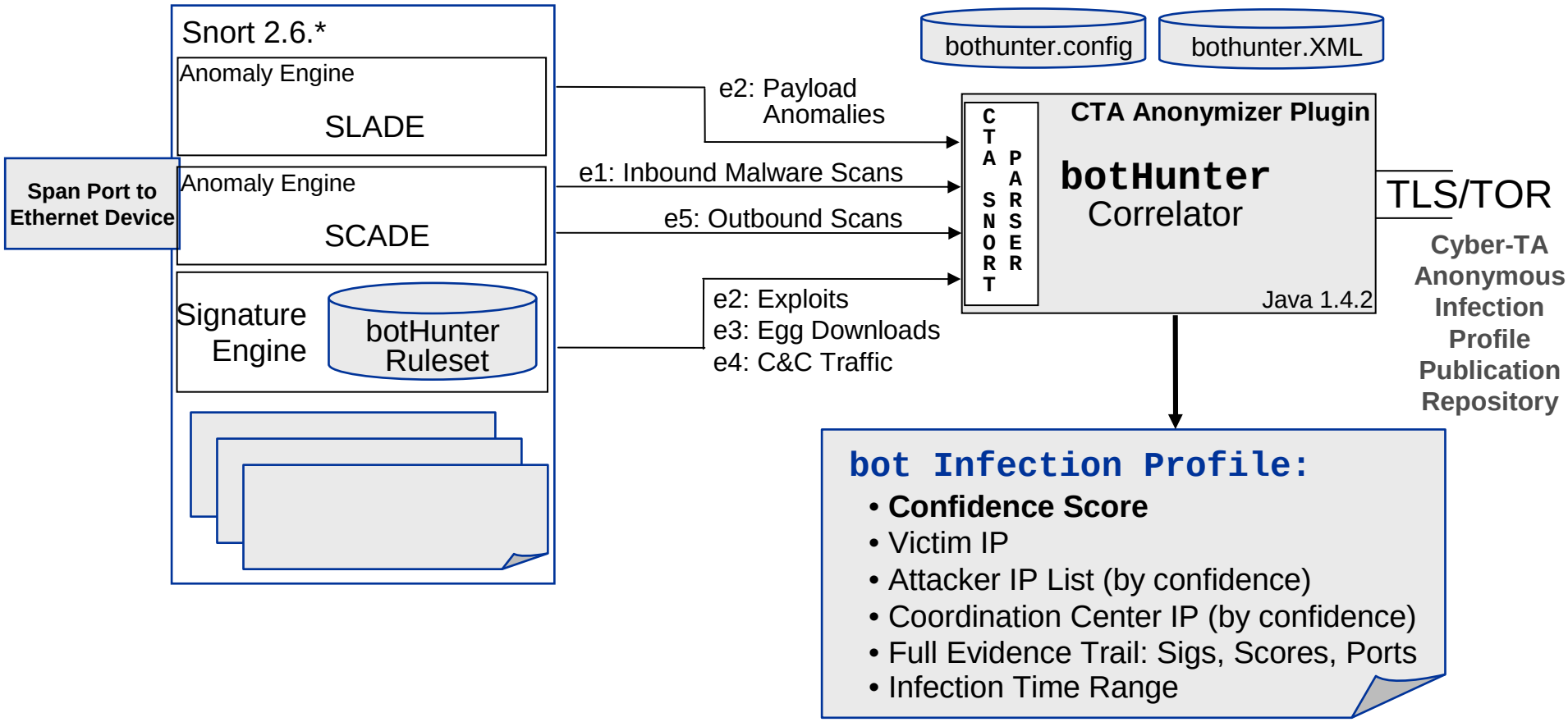
| Int. Host | Timer | E1 ◷ | E2 | E3 | E4 | E5 |
|---|---|---|---|---|---|---|
| 192.168.12.1 | ◷ | $A_a \ldots A_b$ | | | | |
| 192.168.10.45 | ◔ | | $A_c \ldots A_d$ | | $A_e \ldots A_f$ | |
| 192.168.10.66 | ◔ | | $A_g$ | | | |
| 192.168.12.46 | ◔ | | | | $A_h \ldots A_i$ | $A_j \ldots A_k$ |
| : | | | | | | |
| 192.168.11.123 | ◷ ◔ | $A_l$ | $A_m \ldots A_n$ | $A_o$ | | |

**Network Dialog Correlation Matrix**

*Characteristics of Bot Declarations*

- External stimulus alone cannot trigger bot alert
- 2 x internal bot behavior triggers bot alert

# BotHunter: Architecture Overview



# BotHunter Sensor Suite : SCADE

## SCADE: Statistical sCan Anomaly Detection Engine

- Custom malware specific weighted scan detection system for inbound and outbound sources
- Bounded memory usage to the number of inside hosts, less vulnerable to DoS attacks
- Inbound (E1: Initial Scan Phase):
  - suspicious port scan detection using weighted score
  - failed connection to vulnerable port = high weight
  - failed connection to other port = low weight
- Outbound (E5: Victim Outbound Scan):
  - S1 – Scan rate of V over time t
  - S2 – Scan failed connection rate (weighted) of V over t
  - S3 – Scan target entropy (low revisit rate implies bot search) over t
  - Combine model assessments: Or, Majority voting, AND scheme

# BotHunter Sensor Suite : SLADE

## SLADE: Statistical payLoad Anomaly Detection Engine

- Suspicious payload detect: new "lossy" n-gram byte distribution analyzer over a limited set of network services
- Implements a lossy data structure to capture 4-gram hash space: default vector size = 2048. (Versus n=4, $256^4 = 2^{32} \approx$ 4Gb).
- Comparable accuracy as full n-gram scheme: low FP and FN
- General performance comparable to PAYL [Wang2004]: to detect all 18 attacks, the false positive of PAYL is 4.02%, SLADE is 0.3601%

Ke Wang, Salvatore J. Stolfo. "Anomalous Payload-based Network Intrusion Detection", RAID'04

# BotHunter Sensor Suite : Signature Engine

- Signature Set
  - Replaces all standard snort rules with five custom rulesets: e[1-5].rules
- Scope: Dialog content
  - Known worm/bot exploit signatures, shell/code/script exploits, malware update/download, C&C command exchanges, outbound scans
- Rule sources
  - Bleeding Edge malware rulesets
  - Snort community rules
  - Cyber-TA custom bot-specific rules
- Current Set
  - 1383 rules, operating on SRI/CSL and Georgia Tech networks, low FP

# Example BotHunter Infection Profile

Score: 1.95 (>= 0.8)
Infected Target: 192.168.166.40
Infector List: 192.168.166.20
C & C List: 192.168.166.10 (27)
Observed Start: 01/19/2007 17:15:27.60 EST
Report End: 01/19/2007 17:18:26.22 EST
Gen. Time: 01/19/2007 17:18:26.22 EST

**Example VMWare RBot Experiment**
Initial Bot Infector: 192.168.166.20
Victim System: 192.168.166.40
Coordination Center: 192.168.166.10

INBOUND SCAN

EXPLOIT
192.168.166.20 (2) (17:15:27.60 EST)
E2[rb] SHELLCODE x86 0x90 unicode NOOP

EXPLOIT (slade)
192.168.166.20 (2) (17:15:27.60 EST)
E2[sl] Slade detected suspicious payload exploit with anomaly score 2312.725576.

EGG DOWNLOAD
192.168.166.20 (2) (17:15:27.96 EST)
E3[rb] TFTP GET .exe from external source 1028->69 (17:15:27.96 EST)

C and C TRAFFIC
192.168.166.10 (27) (17:15:46.56 EST-17:18:26.22 EST)
E4[rb] BLEEDING-EDGE TROJAN IRC NICK command 1029->6668 (17:15:46.56 EST)
E4[rb] BLEEDING-EDGE TROJAN BOT - potential scan/exploit command
……

OUTBOUND SCAN
192.168.166.20 (17:16:42.18 EST)
E5[sc] scade detected suspicious scanner [192.168.166.40] scanning 30 IPs at ports [0 135 …]

# Detection Performance at SRI Honeynet

| | |
|---|---|
| 3 Month Infection Count | 7,204 |
| Number of Unique Attackers | 6,650 |
| Number of DNS lookups | 2,859 |
| Infections missed by BotHunter | 14 |

**Live Internet Monitoring Apparatus examines BotHunter detection coverage (TP analysis)**

Detection rate > 99.8% so far

**Population Dynamics** | **Network Analyses** | **Host Forensics** | **Binary Analysis**

| Time | Victim OS | Infection Source | C&C Server | DNS Lookups | Infection Port | Packet Trace | Detection Signatures | Infection Chatter | BotHunter Score | BotHunter Profile | Forensic Logs | Antivirus Labels | Packed egg.exe | Unpacked egg.exe | Unpacked egg.asm | Data Strings | Syscall Trace |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 00:26:00 | WinXP | US: 130.13.160.76 | n/a | | 445 | pcap | raw alerts ruleset | ftp 12 lines | Yeah : 0.8 | profile | summary tarball | 10 of 29 | c4709f16a6 | none [4] | none:none | none | trace |
| 00:47:00 | WinXP | TW: 59.104.41.147 | n/a | UA:citi-bank.ru | 445 | pcap | raw alerts ruleset | http 1 line | Yeah : 1.3 | profile | summary tarball | 29 of 29 | d6df3972a0 | 15d31ff96b [0] | ASM:Graph | lines=149 | trace |
| 01:14:00 | Win2K-f | US: 65.141.157.211 | n/a | | 445 | pcap | raw alerts ruleset | http 118 lines | Yeah : 0.8 | profile | summary tarball | 25 of 29 | a7c70c4cbc | a7c70c4cbc [1] | ASM:Graph | lines=697 | trace |
| 02:11:00 | WinXP | US: 71.70.215.119 | n/a | DE:siliconfireware.ru US:ebookfinaltrash.ru :wpad | 445 | pcap | raw alerts ruleset | http 1 line | Yeah : 0.8 | profile | summary tarball | 29 of 29 | a12cab51ef | none [4] | none:none | none | trace |
| 02:46:00 | Win2K-f | CA: 216.167.252.64 | n/a | | 445 | pcap | raw alerts ruleset | http 1 line | Yeah : 0.8 | profile | summary tarball | none | none | none | none | none | none |
| 02:52:00 | WinXP | US: 71.121.157.247 | n/a | UA:citi-bank.ru | 445 | pcap | raw alerts ruleset | http 1 line | Yeah : 1.3 | profile | summary tarball | 29 of 29 | 8ae2cc2e80 | f1ed53cb52 [0] | ASM:none | lines=2 | trace |

http://www.cyber-ta.org/releases/malware-analysis/public/

# Detection Performance at Georgia Tech

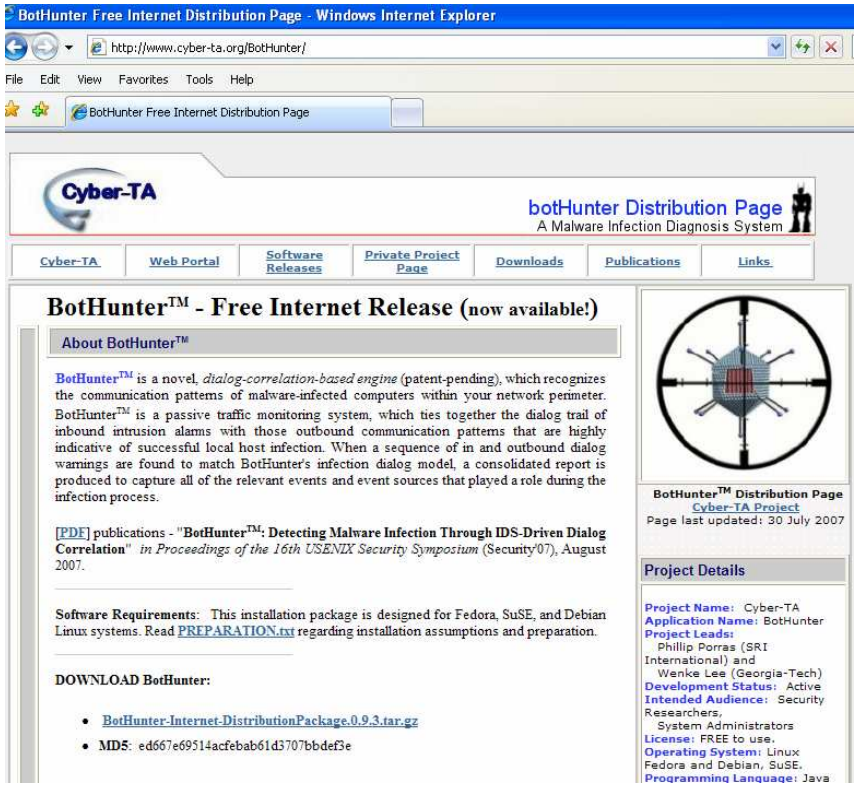
- Virtual network, detect all 10 bots, including
  - AgoBot, Phatbot
  - RBot, RxBot
  - WisdomBot/SdBot/SpyBot
  - GTBot
- Real capture in live network
  - Feb. 2007, Georgia Tech, CoC network
  - BotHunter declared a bot infection via dialog warnings E1, E4, E5
  - E4 (C&C Server) address seen in both Shadow Server and the botnet mailing list

# False Positive Test

- Georgia Tech, college of computing, live deployment
  - Less than 1 (false profiles) per day in a 4 month real-time operation
- SRI computer science lab
  - 1 false positive in a 10-day trace

# Summary

- New network perimeter monitoring strategy: dialog correlation
- New bot detection system: BotHunter
- Free Internet release at




http://www.cyber-ta.org/BotHunter/