Gone in 60 Minutes: Stealing Sensitive Data from Thousands of Systems Simultaneously with OpenDLP

Andrew Gavin
Verizon Business

Presentation Outline

- What is OpenDLP?
- Why write it?
- How does OpenDLP's agent work?
- Benchmarks: Agentless vs agent
- Live demo of agent
- Newly-implemented features
- Future plans
- Q&A

What is OpenDLP?

- A data discovery tool with two components: Agent and web application
- Webapp is LAMP, agent is Windows
- Free and open source (GPLv3)
- Useful for:
  - Compliance personnel
  - Network/System administrators
  - Penetration testers

Why Write It?

- Previous to OpenDLP, there was no free agent-based data discovery tool
- Other available FOSS tools were all designed to be manually run from a single workstation:
  - Cornell Spider (http://www2.cit.cornell.edu/security/tools)
  - FindSSN (Sourceforge)
  - grep
  - These tools could be hacked as agentless scanners (using network shares)
  - Not practical for large deployments

How does OpenDLP work for agent-based scans?

Create a Reusable Policy

- Administrator authentication credentials
  - Can also use pass-the-hash technique instead of password
- Directories and file extensions to whitelist/blacklist
- Memory ceiling for agent (as percent of physical RAM)
- Regular expressions to use (PCREs)
- Concurrent agents to deploy
- Whether to obfuscate sensitive info in database
- How often agents phone home with results

Start a Scan

- Agents deployed over SMB
- Agents started with Samba's "winexe"
- Webapp can concurrently deploy scanners
  - Deploy agents to 1,000 systems in total
  - Can deploy 30 concurrently to make it faster

Agents deploy to Windows systems

- Runs as a service at low CPU priority
- Limits itself to a percent of system memory
- Begins running:
  - Whitelist/blacklists files and directories
  - Begins searching files for regular expressions
  - Securely pushes findings to web server every X seconds
- When done, agent asks to be uninstalled by web application
- Written in C with no .NET requirements

Monitor Agents in Web Application

- Securely receive results every X seconds from agents
  - Current status of agent (directory listing, scanning)
  - How many files it has processed
  - How many bytes it has processed
  - Estimated time to completion
  - Two-way-trusted SSL connection
- Can pause or uninstall agents at any time
- Automatically deletes and uninstalls agents when done

Review Results in Web Application

- View high level information about entire scans
  - Each scanner's number of findings
  - Each scanner's estimated time of completion
- View detailed information about specific scans
  - Findings with filenames, byte offsets
  - Hyperlinks to download files with findings:

So wait a minute…

You invented multiplayer grep?

Agent vs Agentless Benchmarks

OpenDLP agent's system's specs:

- Core2duo P8600 (2.4 GHz)
- 4 GB RAM
- 7200 RPM, 250 GB HDD
- 100 mbit network

Benchmark: OpenDLP Agent

OpenDLP run time on one system

- 13 regexes scanned 2.05 GB in 01:07:39
  - 04:15 to enumerate/blacklist files, read files into memory
  - 01:03:24 to perform calculations
  - (Negligible time to install/uninstall agent, upload results)
  - 1 GB scanned every 32:57 with 13 regexes
- Extrapolation: With just one regex = 09:07
  - 04:15 to enumerate/blacklist files, read files into memory
  - 04:52 to perform calculations
  - 1 GB scanned every 04:45

Benchmark: Agentless

Agentless scanner's run time for one system

- 13 regexes scanned 2.05 GB in 01:20:26
  - 17:02 to download/read all files
  - 01:03:24 to perform calculations
  - 1 GB scanned every 39:10
- Extrapolation: With just one regex = 21:54
  - 17:02 to download/read files
  - 4:52 to perform calculations
  - 1 GB scanned every 10:40

Benchmark Comparison

Agent-based vs. agentless for one system

- 13 regexes: Agentless 19% slower
- 1 regex: Agentless is 130% slower
- For one system, performance hit might be worth not installing agent

What if we extrapolate this to more systems?

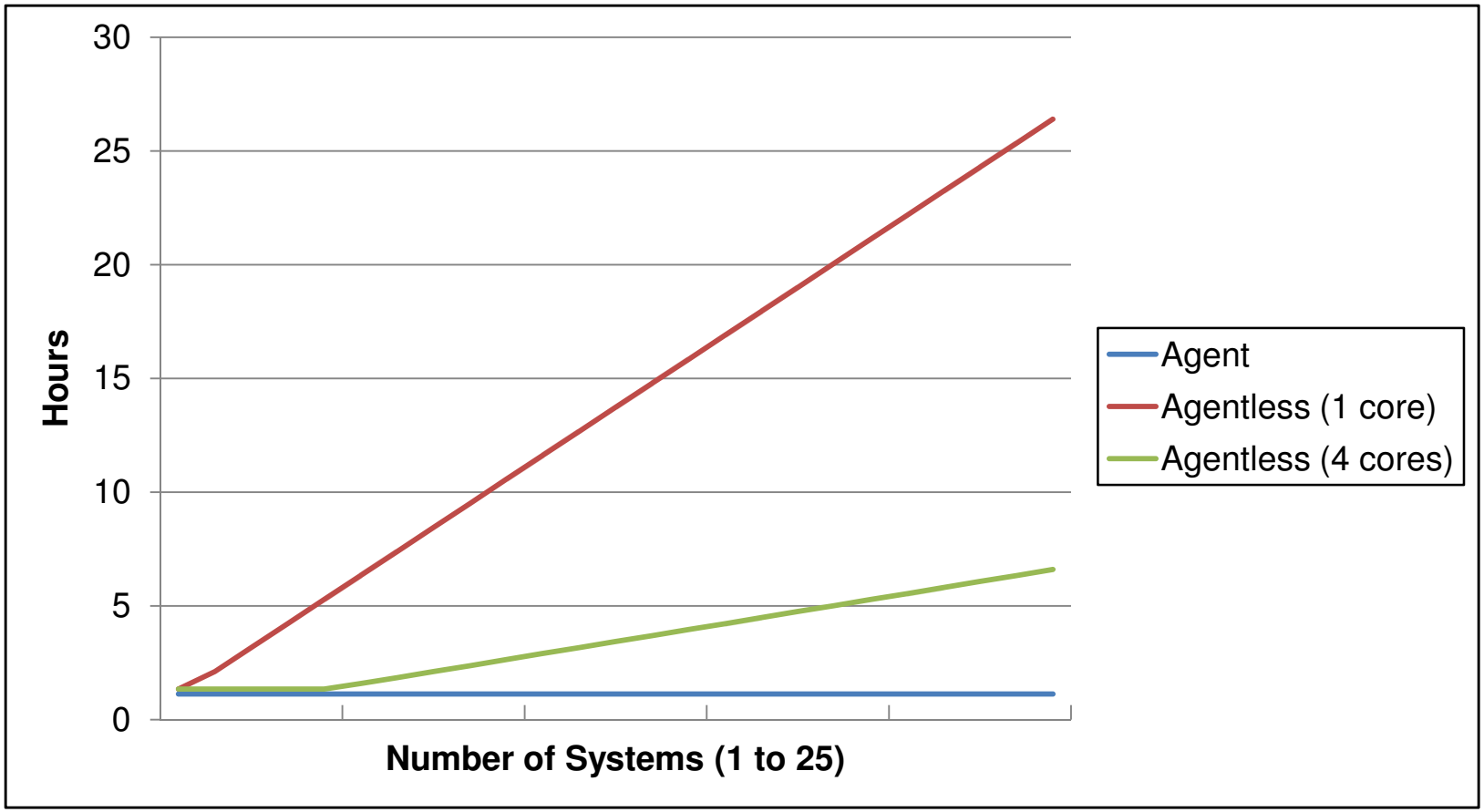
Benchmark: Agentless Bottlenecks

Agentless with 13 regexes: 01:20:26

- Network (100 mbit): 17:02 wallclock (21.2%)
  - 16.5 mbit throughput over SMB (directory crawling and file downloading)
  - On 100 mbit network, can do 6.06 systems concurrently without bottleneck
- CPU: 01:03:24 wallclock (78.8%)
  - On single core, can do 1.27 systems concurrently
  - On quad core, can do 5.08 systems concurrently

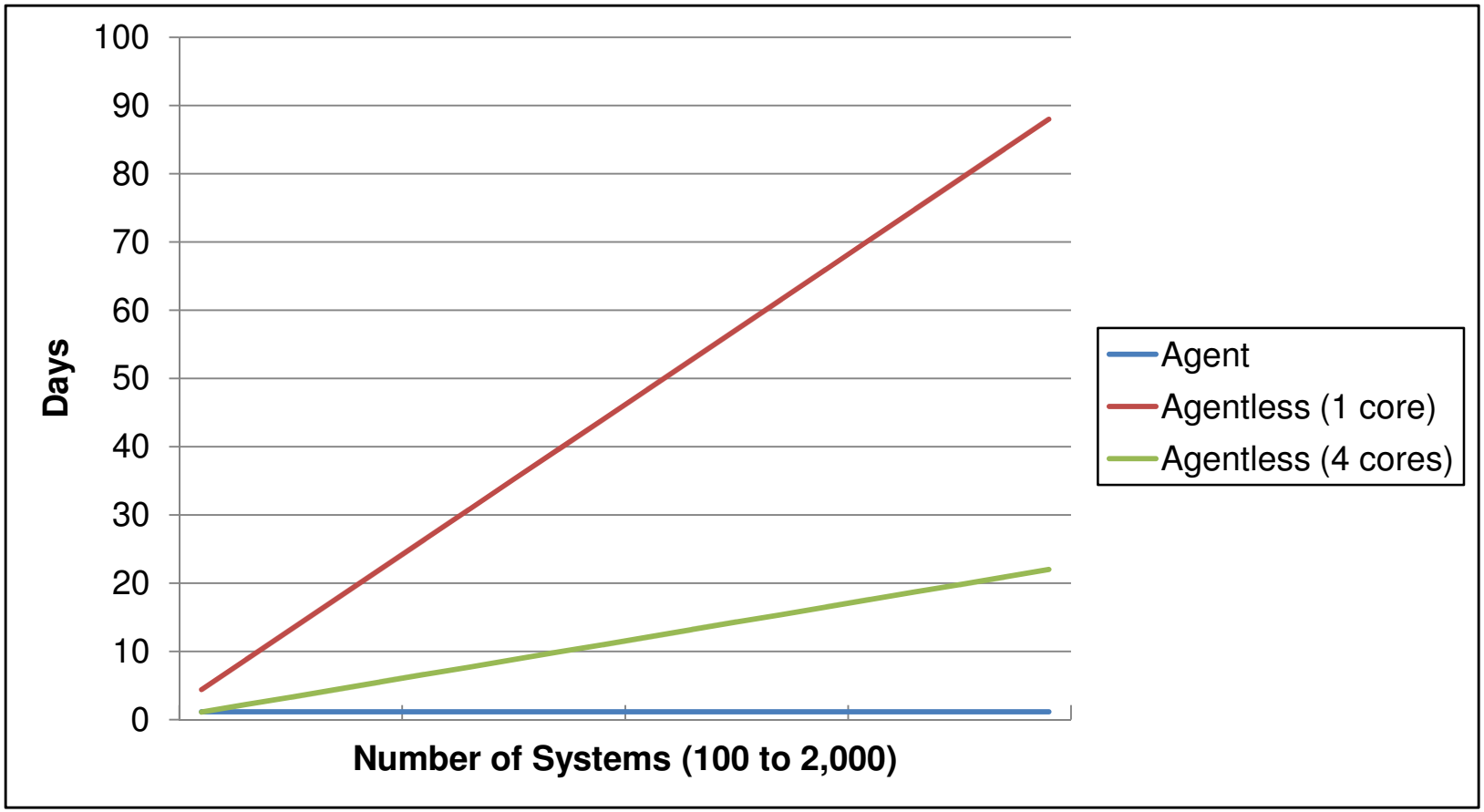
Benchmark: 1 to 25 systems

Agent vs Agentless Time Comparison



Benchmark: 100 to 2,000 systems

Agent vs Agentless Time Comparison



Agent vs Agentless Benchmark Results

Agent-based upsides:

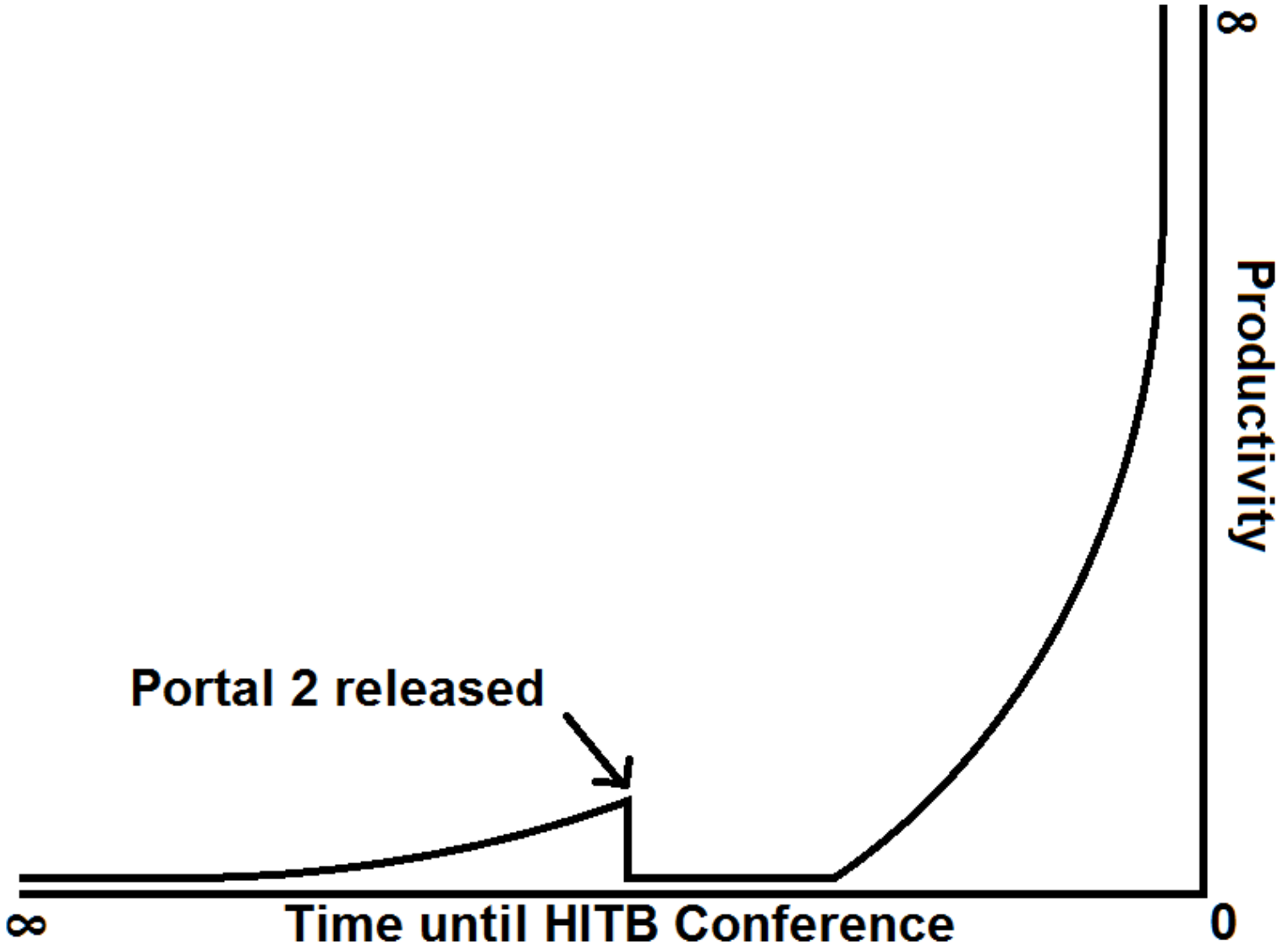
- All computations distributed to victim systems
- Minimal network traffic
  - OpenDLP agent is only 1.02 MB compressed
  - Only logs and results uploaded to webapp

Agentless downsides:

- All computations done on central system
- All files must be downloaded (over SMB) to central system

Live Demo of Agent Scan

New features



Agentless database scans

Create a Reusable Policy

- Database authentication credentials ("sa", "root", etc)
- Databases, tables, and columns to whitelist/blacklist
- Number of rows to grab (or grab all rows)
- Regular expressions to use (PCREs)
- Concurrent scans to deploy on your operating system
- Whether to obfuscate sensitive info in OpenDLP database

Start a Scan

- Scans started on your own system in background as Perl script
- Webapp can concurrently start scanners on several database servers
- Will traverse database structure just like what you do with SQL injection:
  - Grab database names, table names, column names
  - Go after actual data

Supported Databases

- Microsoft SQL server (using FreeTDS driver and DBI Perl module)
  - Sybase?
  - Microsoft SQL Server Express?
- MySQL (using DBI Perl module)

- Databases are very easy to add
- More will be coming in near future (DB2, Oracle, PostgreSQL)

Live Demo of Agentless Database Scan

Future Plans

- Agentless option (50% done)
- Scan more databases
- More agents (Linux, OSX)
- Output in Word/Excel
- Trending graphs (Excel/ImageMagick)
- Metasploit integration
- Portable agent (deploy on USB thumbdrive for use during social engineering attacks)
- ~~Monitor PCs for network traffic and file copying~~ - See MyDLP project (www.mydlp.org)

Availability, Contact Info, Q&A

- http://opendlp.googlecode.com
  - 0.3.1 source code and binaries
  - 0.2.2 Ubuntu-based VirtualBox VM
- andrew.opendlp@gmail.com
- http://twitter.com/OpenDLP

- Q&A