

---

# Poor Man's Data Leak Prevention

---

Chris Brenton  
cbrenton@chrisbrenton.org

---

# Where Can I Get A Copy Of These Slides?

---

[www.chrisbrenton.org](http://www.chrisbrenton.org)

---

# What We Will Cover

---

- Open source Vs. commercial
- DLP at the host level
- DLP at the network level
- Additional steps you can take

---

# Is It Cost Effective To Build DLP Yourself?

---

- Can't match a complete commercial host/network solution
  - Yet...
- But you can plug some of the holes
- Expectations
  - Can check for leakage on the wire
  - Can scan files for signatures
  - Can real time check for SSN and CCN
  - Real time custom checks still a work in progress
  - Product options limited

---

# Where To Begin

---

- Same as you would for any DLP deployment
- Info discovery similar
- Consolidation is similar
- Architecture changes are similar
- What's different
  - More puzzle pieces
  - Less integration
  - Expect to glue together many different tools
  - Expect to do a lot of scripting

---

# Inventory

---

- What type of data do you need to protect?
  - Social Security Numbers (SSN)
  - Credit Card Numbers (CCN)
  - Salaries
  - Corporate books
  - Future products or projects
  - Contact information
- Talk to the people handling the data
- If unsure if it is sensitive, include it

---

# Numbers To Look For

---

- SSN format

[http://www.security.vt.edu/Downloads/legit\\_ssn.pdf](http://www.security.vt.edu/Downloads/legit_ssn.pdf)

- Visa

[http://www.security.vt.edu/Downloads/legit\\_visa.pdf](http://www.security.vt.edu/Downloads/legit_visa.pdf)

- MasterCard

[http://www.security.vt.edu/Downloads/legit\\_mc.pdf](http://www.security.vt.edu/Downloads/legit_mc.pdf)

- American Express

[http://www.security.vt.edu/Downloads/legit\\_amex.pdf](http://www.security.vt.edu/Downloads/legit_amex.pdf)

---

# What Else To Look For

---

- Look for multiple SSN or CCN matches
  - Reduces false positive rate
- Multiple dollar signs in a single document
- Keywords such as “company private”
  - Do you tag documents?
- Possible zip codes
- Use your imagination

---

# Discovery

---

- Plan on checking
  - File and Web servers
  - Desktops
  - Laptops
  - USB inventory
  - ???
- Determine what tools you will need
- Define a process for discovery

# Checking File Shares

- Find\_SSNs from Virginia Tech
- Supports
  - Linux/UNIX (Python)
  - Windows (pre-compiled binary)
- Searches for
  - US SSNs
  - Visa, MC and AmEX numbers
- Safeguard log output!

<http://security.vt.edu/findssnccn.html>

---

# More File Tools

---

- Sensitive number finder (SENF)
  - University of Texas
- Searches for SSNs and CCNs
- Java based tool
- Makes a nice zero footprint client
- Redirect outbound Web sessions
  - User would need write access to log

<https://senf.security.utexas.edu/wiki/WikiStart>

---

# Spider For Crawling

---

- Cornell University Spider
- Supports Win, Mac, Linux, UNIX
- Use to crawl internal & external Web sites
- Shows what is anonymously accessible
  - Will not show what is accessible via SQL magick
- Specifically checks for SSNs and CCNs

<http://www2.cit.cornell.edu/security/tools/>

---

# Beyond SSNs And CCNs

---

- Any RegX compatible tool
- Some free time for scripting
- Possible tools
  - grep
  - Awk & Sed
  - Perl or Python
  - All require Cygwin for Windows
- Native tools
  - “find”, but it is limited

---

# Got The Data, Now What

---

- Figure out what's needed, what's not
- Isolate sensitive data from non-sensitive
- Reduce the number of data stores
  - Minimizes what needs to be protected
- Divide the info into different stores
  - Minimizes quantity of info per file
- Leverage network drives for access control
  - Also auditing and logging
- Client/Server to isolate users from data

---

# Set Proper Permissions

---

- Leverage Folder and file permissions
- Restrict access as much as possible
- Be careful with database front ends
  - Front ends can be circumvented
- Permissions will not prevent data leaks
  - Just minimize the conduits for leakage

---

# Audit File Access

---

- Audit trail of who gained access to the info
- Will not prevent data leaks
  - Forensics trail if the worst occurs
- Enable auditing
  - Admin Tools → Security Policy → Audit Policy → Audit Object Access
- Specify what to audit
  - Right click file → Properties → Security → Auditing

---

# Leverage DRM When Possible

---

- Microsoft Digital Rights Management
- Restrict access to MS Office files
- Define restrictions at the user level
- Can restrict copy, print screen functions
- Can set content to expire
- Still susceptible to:
  - keystroke loggers and other Malware
  - 3<sup>rd</sup> part screen captures
  - Cell phone cameras or pen and paper
  - Been cracked more than once already

---

# Checking Local Systems

---

- ClamAV
  - Open source
  - anti-virus, phishing, checks
- Windows, Linux, UNIX agents
- Also integrates with many mail servers
- Libclamav has modular support for DLP
  - SSN and CCN checking
  - Signature customization in the works

---

# Next Step: The Network

---

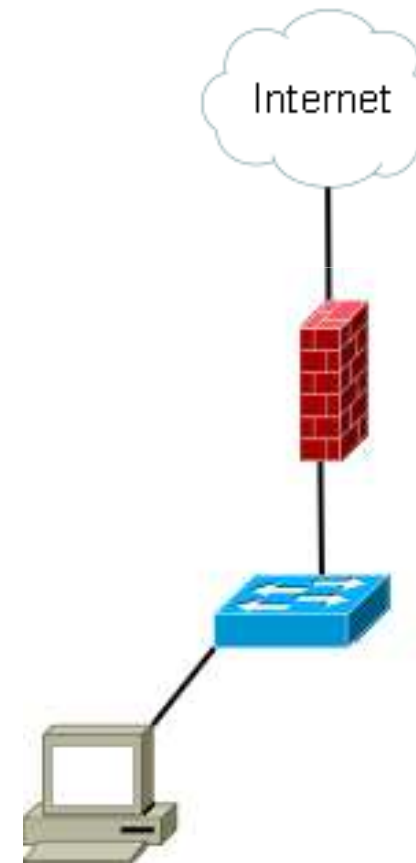
- Evaluate your current network design
- Is end to end encryption supported?
- Any choke points where content can be monitored?
- Can we proxy e-mail?
- Can we proxy HTTP?
- Can we enforce site restrictions?

---

# End To End Encryption

---

If users can encrypt their sessions, the host is the only place you can perform content checking



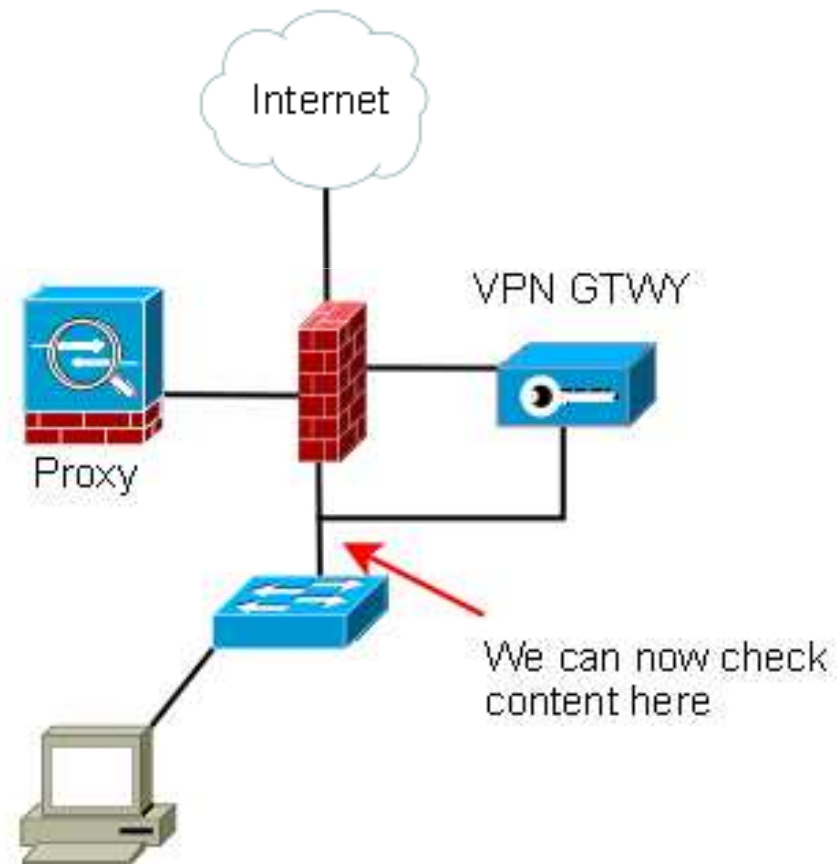
---

# Questions To Ask

---

- What job roles have a business need for end to end encryption?
  - If any
- What to evaluate
  - Inbound/Outbound IPsec
  - SSL/TLS of TCP applications
  - Inbound/Outbound SSH (via port forwarding)
- Can we leverage an outbound proxy?
  - Permits content checking + encryption

# Checking Content For Leaks



---

# HTTP Proxy With Squid

---

- Can combine with SquidGuard and ClamAV
- Squid does not support DLP
  - ClamAV does as a plug-in
- The real benefit is:
  - Prevention of end to end encryption
  - User level access control of Web sites
  - Filtering out "PUT" solves a lot of problems
    - But may create others

<http://www.squid-cache.org/>

---

# SMTP Proxy

---

- Spamassassin
  - One of the top spam detection tools
- Run a second process for DLP checks
  - Requires heavy .cf file modification
- “Score” the likelihood of being a DLP event
- Leverage MTA to quarantine
- Other good reasons, covered in another slide

<http://spamassassin.apache.org/>

# Simple Content Checking With Netfilter

```
iptables -A FORWARD -p tcp --tcp-flags ACK ACK -m string --string "Financial Summary"  
-j LOG --log-prefix " Possible Data Leak "  
iptables -A FORWARD -p tcp --tcp-flags ACK ACK -m string --string "Financial Summary"  
-j REJECT --reject-with tcp-reset
```

It is simple to setup, but that simplicity comes at a cost

# More Advanced Content Filtering

- Snort + Netfilter

```
iptables -I FORWARD -j QUEUE
```

- Can run as an IDS or IPS
- Adds RegX capability
- Existing Snort rules for detecting conduits
  - Chat, IRC, etc.
- No default rules for DLP... yet

<http://www.snort.org/>

---

# Ngrep

---

- “Grep” for network packets
- Versions for Windows, Linux, UNIX
- Supports RegX capability
- Save suspect traffic as Libpcap
- Load patterns to match from a file (-F)
  - Permits logical AND/OR’s
  - Can define fairly complex filtering

<http://ngrep.sourceforge.net/>

---

# Phishing & Malware

---

- Majority of info leaks are stolen info
- Protecting SMTP
  - MailScanner
  - SpamAssassin
  - ClamAV
- Protecting HTTP
  - Squid
  - SquidGuard
    - Block known Malware sites

---

# Tighten Outbound Access

---

- Evaluate your needs for outbound access
- The dreaded “ANY” outbound rule
  - Should Malware really be allowed to TFTP down its rootkit?
  - From the financial server?
- Limit outbound access to what is needed
- Log ***everything*** you permit out
- Define security zones and segregate

---

# Monitor Your Logs

---

- If you are auditing access to critical files, centrally log the info
- This can permit you to setup a real time alert system for DLP events
- Swatch
  - Simple to learn but limited in abilities
- Simple Event Coordinator (SEC)
  - Extremely flexible but steep learning curve

<http://sourceforge.net/projects/swatch/>

<http://www.estpak.ee/%7Eristo/sec/>

---

# Google As Your Friend or Enemy

---

- Leverage Google to crawl your site
- Before the bad guys do it for you
- Queries can expose
  - Sensitive files or directories
  - Mis-configurations
  - Vulnerabilities
  - Revealing error messages or codes

<http://www.ethicalhacker.net/content/view/41/2/>

---

# Backup Logs

---

- Detailed logs will show every file backed up
- Will also reveal when it was stored on the system
- This can help to locate and track sensitive files across the network
- GUI log tool can make queries cumbersome
- Command line is quickest
  - SQL queries
  - Grep or Find searches

---

# Final Word

---

- Doing DLP on the cheap requires a lot of pieces
- You need to weigh time investment Vs. the cost of a dedicated product
- You may have noticed that many of the steps we discussed
  - Also address Malware concerns
  - Improve security posture
  - Improve overall network awareness

---

# Questions?

---

- Where can I get the slides?
  - [www.chrisbrenton.org](http://www.chrisbrenton.org)
- Contact info
  - [chris@chrisbrenton.org](mailto:chris@chrisbrenton.org)