Using certificate transparency streams to hunt down phishing sites

Sean Gallagher
@thepacketrat
IT/NatSec Editor, Ars Technica
Certificate data, and why it's great for OSINT

- More phishing sites are using SSL certs, other tools to evade filters
- LetsEncrypt is free, programmatic, as is cPanel – which auto-cert registers new domains
- While DNS holds domain names of sites, don't identify malicious domains in DNS until they go active.
- Phishing domains may have already switched by the time activity is detected.
- Certificate data can be scanned actively to discover phishing sites before they go live, and used to proactively collect forensic data on kits, etc. while the sites are still being set up.
Certificate Transparency Logs

- Were created to help track the validity of SSL certificates issued by the major certificate authorities.
- Data helps detect mistakenly issued certificates, stolen or otherwise maliciously acquired certs, among other things.
- Each CA logs each new certificate, flags revocations, etc.
The "Certstream"

- Service created by Ryan Sears of Cali Dog Security that generates a near-realtime data feed from all of the major certificate authorities' transparency logs.
- Data is published over a web socket, can be directly accessed via a command line interface published in pip (pip install certstream)
- Gives you a timestamp for an event, the source url for the log it came from, and the domain associated with the certificate.
- This by itself, with a little grepping with regular expressions, or minor coding, could be used to hunt for threats spoofing specific domain name patterns.
Censys.io stores certificate log data so it can be used for searching against recent and historical certificate registration.
Targeted hunting in CTL: search by pattern
Targeted hunting in CTL: certificate history
Targeted hunting in CTL: whois
Targeted hunting in CTL: browsing/capture
Using ML to scan: StreamingPhish

```
rodan-3:streamingphish admin$ docker-compose exec cli streamingphish

by Wes Connell
@wesleyreptor

1. Deploy phishing classifier against certstream feed.
2. Operate phishing classifier in manual mode.
3. Manage classifiers (list active classifier and show available classifiers).
4. Train a new classifier.
5. Print configuration.

Please make a selection (1-6): 4
[*] Loading benign data.
[*] Loading malicious data.
[*] Completed loading training data.
[*] Computing features...
[*] Training complete.
[*] Computing classifier metrics...
[*] Classifier metrics available.
```
StreamingPhish

- StreamingPhish: https://github.com/wesleyraptor/streamingphish
- Wes Connell – works in threat detection at Uber-developed this ML tool
- Uses training sets of data, including brands and TLDs, and editable set of malicious domains- to build an algorithm to spot potentially malicious domains as they appear in certstream.
- Docker based, can run in cloud
StreamingPhish

CLI output
Next steps and further automation

- Censys Whois dig, DomainTools, SecurityTrails, ViewDNSinfo
  - Neighboring IPs
  - Additional domains at host
- Kit and shell hunting
- Lookout is automating the rest of the kill chain with ML – download kits, analyze screens, block
Digging for phish: Censys Whois Dig
Digging for phish: SecurityTrails
Digging for phishing: SecurityTrails
Digging for phish: ViewDNS.info
Questions?

- Email: sean.gallagher@arstechnica.com
- Sample kits found: https://github.com/packetrat/phishing-kits-I-found
- Twitter: @thepacketrat