PAPER PRESENTATION: HIGHLY PREDICTIVE BLACKLISTING

John Bambenek

CS 563

PROBLEM

- There are “tons” of malicious events detected by firewalls, intrusion detection systems, web application firewalls, etc.
- The adversarial infrastructure may be persistent, may be a VPS, compromised host, etc.
- Can I determine both what is most relevant to my organization and relevant globally that will be worth blocking “in the future”?

PROBLEM

- Consider your typical firewall:

- iptables –A INPUT –p 80 –j ACCEPT
  - What does this not protect against?

WHAT IS DSHIELD?

- Run by SANS (I'm one of the Handlers) where people submit firewall and IDS block logs from around the world.
- Also can operate a DShield sensor as a raspberry pi. Primarily finds port-level blocks and darknet traffic.
- Each user has their own ID, can also "action" blocks. In turn, this gives a huge dataset that is "mostly" globally representative about "loud attacks".

THREE APPROACHES

- Global Worst Offender Lists (GWOL)
  - Misses targeted or localized attacks
- Local Worst Offender Lists (LWOL)
  - Misses attacks that may not have "gotten there" yet

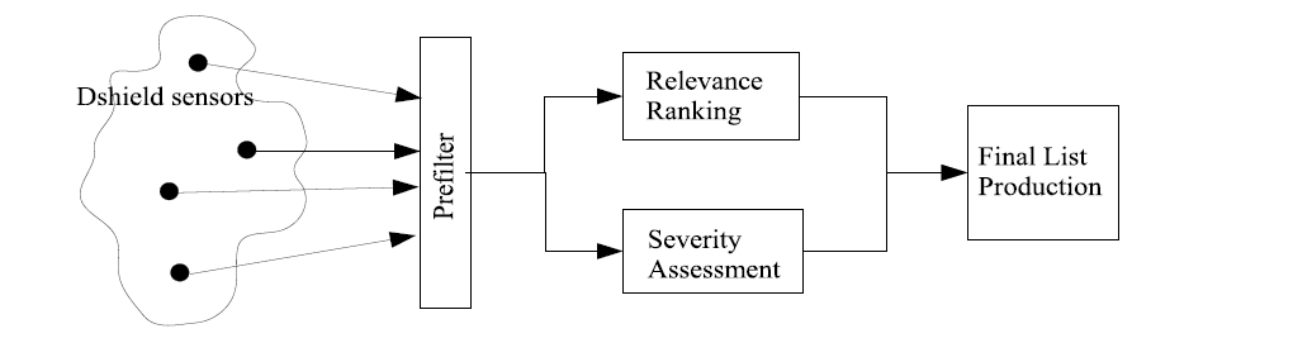
- This paper introduces Highly-Predictive Blacklist (HPB) that uses elements of both.

HPB APPROACH

- Analogous to Google PageRank

- Incorporates the following:
  - Log prefiltering (i.e. RFC 1918 addresses, "local" addresses, etc
  - Relevance based ranking (per-contributor basis)
  - Severity analysis (looks at known malware propagation patterns)

ARCHITECTURE



PRE-FILTERING

- Drop the obvious noise:
  - RFC 1918 addresses
  - Bogons
  - Unassigned IPs
  - Why?
- Drop “internet measurement” services, crawlers, etc. Why?
- Drop common ports (80, 53, 25, 443)

RELEVANCE RANKING

- How "close" is a specific attacker to a specific victim?
  - If you have enough data about many victims, you can see patterns and order of how attacks progress through internet. (i.e. Attacker X will always hit Victim A 2 days before Victim B.)

	v_1	v_2	v_3	v_4	v_5
s_1	*	*			
s_2	*	*			
s_3	*		*		
s_4		*	*		
s_5		*			
s_6				*	*
s_7			*		
s_8			*	*	

Table 1: Sample Attack Table

RELEVANCE RANKING

- Create a matrix based on (m_{ij} / m_i) (common attack sources / all attack sources) for each relationship between victims and sources. (First pass)

$$\begin{pmatrix} 0 & 0.33 & 0.083 & 0 & 0 \\ 0.33 & 0 & 0.063 & 0 & 0 \\ 0.083 & 0.063 & 0 & 0.13 & 0 \\ 0 & 0 & 0.13 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0 \end{pmatrix}$$

Figure 2: Standardized Correlation Matrix for Attack Table 1

- $R^s = W \times b^s$ (Relvancy vector is product of Adjacency matrix and attack vector)

RELEVANCE WITH "LOOK AHEAD"

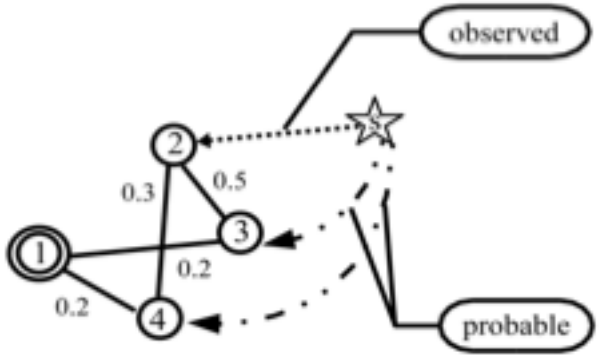


Figure 3: Relevance Evaluation Considers Possible Future Attacks

PROPAGATING RELEVANCY

- Better version is:

$$\mathbf{r}^s = \sum_{i=1}^{\infty} (\alpha \mathbf{W})^i \cdot \mathbf{b}^s$$

- Solving for x:

$$\mathbf{x} = \mathbf{b}^s + \alpha \mathbf{W} \cdot \mathbf{x}$$

- This gives something used by PageRank to figure relevant results.

ATTACK SEVERITY

- Note: This paper was done in 2008. This is important.
- Malicious behavior modeled after typical "scan-and-infect" behavior.
- Calculates based on /24 network basis.
- Three factors used: Port Score, Target Count, International Victim Count

$53 - UDP$	$69 - UDP$	$137 - UDP$	$21 - TCP$	$53 - TCP$	$42 - TCP$
$135 - TCP$	$139 - TCP$	$445 - TCP$	$559 - TCP$	$1025 - TCP$	$1433 - TCP$
$2082 - TCP$	$2100 - TCP$	$2283 - TCP$	$2535 - TCP$	$2745 - TCP$	$2535 - TCP$
$3127 - TCP$	$3128 - TCP$	$3306 - TCP$	$3410 - TCP$	$5000 - TCP$	$5554 - TCP$
$6101 - TCP$	$6129 - TCP$	$8866 - TCP$	$9898 - TCP$	$10000 - TCP$	$10080 - TCP$
$12345 - TCP$	$11768 - TCP$	$15118 - TCP$	$17300 - TCP$	$27374 - TCP$	$65506 - TCP$
$4444 - TCP$	$9995 - TCP$	$9996 - TCP$	$17300 - TCP$	$3140 - TCP$	$9033 - TCP$
$1434 - UDP$					

Figure 5: Malware Associated Ports

LIST PRODUCTION

- Then just sort by score and pick X to generate the list.
  - All protective technologies (firewalls, routers, etc) have limits in how many entries they can accept.
- Results showed a 20-30% increase.

	Increase Average	Increase Median	Increase StdDev	Increase Range
vs. GWOL	129	78	124	40 to 732
vs. LWOL	183	188	93	59 to 491

Table 5: Top 200 Contributors' Hit Count Increases (Blacklist Length 1000)

RISKS

- Can a false positive entry be included?
  - There is a global white-list but not a localized one (and more importantly, there is no "good" global whitelist. (Some of my upcoming research).
- Can an attacker get their attacks excluded?
  - Can be a sensor and try to break various elements of alignment but requires broad (but not complete) knowledge of the ecosystem and relationships.
- Can all the data be poisoned?
  - It's a volunteer system, so anyone can join and dump in junk data

CURRENT STATE

(Not in paper)

- SRI has ”abandoned” the code.
- DShield no longer generates HBPLs.
- *Incoming* attack data is not as important as *outgoing* attack data.
  - Malware beacons out now, reverse shells are common. Best way to beat a firewall is to have a machine on inside using existing ACLs.

QUESTIONS?