Behavioral Malware Detection in Delay Tolerant Networks

Wei Peng, *Student Member, IEEE,* {Feng Li, Xukai Zou}, *Member, IEEE,* and Jie Wu, *Fellow, IEEE*

1 Design Details

1.1 Posterior $P(S_j|\mathcal{A})$

We have the following observations:

- By the *principle of maximal entropy* [1] (which states that, subject to known constraints, or *testable information*, the probability assignment that best represents our state of knowledge is the one which maximizes the *entropy*, as defined by Shannon [2]), before obtaining any assessment, a node i, which *holds no presumption on another node j's suspiciousness*, should assign a *uniform* distribution to the prior $P(S_j)$, which is:

$$P(S_j) = 1, \tag{1}$$

  since, by definition, $S_j \in [0,1]$. Any other assignment of $P(S_j)$ reflects prejudice that i holds against j, which is *not* warranted by our assumption on the background knowledge B.
- The independence between pairs of assessments implies the *equivalence* of *batch* and *sequential* computation for $P(S_j|\mathcal{A})$. If we apply the assessment sequentially by using the posterior of the previous round as the prior of this round, we have:

$$\begin{aligned}
P(S_j|\mathcal{A}) &= P(S_j|a_1,\ldots,a_A) \\
&\propto P(a_D|S_j,a_1,\ldots,a_{D-1}) \\
&\quad \times P(S_j|a_1,\ldots,a_{A-1}) \\
&= P(a_D|S_j) \times P(Sj|a_1,\ldots,a_{A-1}) \\
&\ldots \\
&\propto P(S_j)\prod_{k=1}^{D} P(a_k|S_j).
\end{aligned} \tag{2}$$

By the definition of suspiciousness S_j and the independence among assessments, we have:

$$P(a_k|S_j) = \begin{cases} S_j & \text{for} \quad a_k = 1 \\ 1 - S_j & \text{for} \quad a_k = 0 \end{cases}. \tag{3}$$

By Equations 1, 2, and 3, we have:

$$P(S_j|\mathcal{A}) \propto S_j^{s_\mathcal{A}}(1-S_j)^{A-s_\mathcal{A}},$$

in which $s_\mathcal{A}$ is the number of suspicious assessments in $\mathcal{A}$ (i.e., the assessments equal to 1), and $A = |\mathcal{A}|$ is the number of assessments collected so far.

1.2 Posterior Maximizer

We can calculate the $S_j \in [0,1]$ which maximizes $P(S_j|\mathcal{A})$. Let $a = s_\mathcal{A}$ and $b = A - s_\mathcal{A}$. If $a = 0$ and $b \neq 0$, $S_j = 0$ is the maximizer; conversely, if $a \neq 0$ and $b = 0$, $S_j = 1$ is the maximizer. If both a and b are both non-zero, let $\mathcal{C}$ be the normalization constant (which is a constant for S_j), we have:

$$\begin{aligned}
\frac{\mathrm{d}P(S_j|A)}{\mathrm{d}S_j} &= \frac{\mathrm{d}}{\mathrm{d}S_j}\left(\mathcal{C}S_j^a\sum_{k=0}^{b}\binom{b}{k}(-S_j)^k\right) \\
&= \mathcal{C}aS_j^{a-1}\sum_{k=0}^{b}\binom{b}{k}(-S_j)^k \\
&\quad - \mathcal{C}bS_j^a\sum_{k=0}^{b-1}\binom{b-1}{k}(-S_j)^k \\
&= \mathcal{C}S_j^{a-1}(1-S_j)^{b-1}\left(a(1-S_j) - bS_j\right).
\end{aligned}$$

The unique $S \in (0,1)$ which makes $\frac{\mathrm{d}}{\mathrm{d}S_j}P(S_j|A) = 0$ is the S_j which satisfies $a(1-S_j) - bS_j = 0$, i.e., $S_j = \frac{a}{a+b}$. Moreover, it maximizes $P(S_j|\mathcal{A})$, even when either a or b (but not both) is zero. Therefore, we have:

$$\operatorname*{arg\,max}_{S_j\in[0,1],\mathcal{A}\neq\emptyset} P(S_j|\mathcal{A}) = \frac{a}{a+b} = \frac{s_\mathcal{A}}{A}.$$

1.3 Monotonicity of $P_g(\mathcal{A})$ and $P_e(\mathcal{A})$ on $s_\mathcal{A}$

We have $P_g(\mathcal{A}) = 1 - P_e(\mathcal{A})$. Thus, we only need to prove the monotonicity of any one of them; the other follows naturally.

Here, we prove that $P_g(\mathcal{A})$ is a monotonically decreasing function on $s_\mathcal{A}$.

- *W. Peng and Dr. X. Zou are with the Department of Computer and Information Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN, 46202.*
- *Dr. F. Li is with the Department of Computer, Information, and Technology, Indiana University-Purdue University Indianapolis, Indianapolis, IN, 46202.*
- *Dr. J. Wu is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA, 19122.*

Let $a = s_{\mathcal{A}}$ and $b = A - s_{\mathcal{A}}$; we only need to prove:

$$(\int_0^1 S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j)^{-1}\int_0^{L_e} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j$$
$$\geq (\int_0^1 S_j^{a+1}(1-S_j)^{b}\,\mathrm{d}S_j)^{-1}\int_0^{L_e} S_j^{a+1}(1-S_j)^{b}\,\mathrm{d}S_j,$$

or, equivalently:

$$\int_0^1 S_j^{a+1}(1-S_j)^b\,\mathrm{d}S_j\int_0^{L_e} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j$$
$$\geq \int_0^1 S_j^{a}(1-S_j)^{b+1}\,\mathrm{d}S_j\int_0^{L_e} S_j^{a+1}(1-S_j)^{b}\,\mathrm{d}S_j.$$

Subtract $\int_0^{L_e} S_j^{a+1}(1-S_j)^b\,\mathrm{d}S_j\int_0^{L_e} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j$ from both sides, we get:

$$\int_{L_e}^1 S_j^{a+1}(1-S_j)^b\,\mathrm{d}S_j\int_0^{L_e} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j$$

for the left side and:

$$\int_0^{L_e} S_j^{a+1}(1-S_j)^b\,\mathrm{d}S_j\int_{L_e}^{1} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j$$

for the right side.

Finally, we have:

$$\text{left} = \int_{L_e}^1 S_j^{a+1}(1-S_j)^b\,\mathrm{d}S_j\int_0^{L_e} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j$$
$$\geq \int_{L_e}^1 L_e S_j^{a}(1-S_j)^b\,\mathrm{d}S_j\int_0^{L_e} (1-L_e)S_j^a(1-S_j)^{b}\,\mathrm{d}S_j$$
$$= \int_0^{L_e} L_e S_j^{a}(1-S_j)^b\,\mathrm{d}S_j\int_{L_e}^{1} (1-L_e)S_j^a(1-S_j)^{b}\,\mathrm{d}S_j$$
$$\geq \int_0^{L_e} S_j^{a+1}(1-S_j)^b\,\mathrm{d}S_j\int_{L_e}^{1} S_j^a(1-S_j)^{b+1}\,\mathrm{d}S_j = \text{right}.$$

Thus, we have proven that "$P_g(\mathcal{A})$ is a monotonically decreasing function on $s_{\mathcal{A}}$" and "$P_e(\mathcal{A})$ is a monotonically increasing function on $s_{\mathcal{A}}$".

2 How to choose the lookahead λ

In this section, we discuss how to adapt the look-ahead λ to individual nodes' intrinsic risk inclinations against the malware.

λ must be large enough so that the decision process will not terminate prematurely. For example, after the first suspicious-action assessment against J, depending on L_e, the evidence might become unfavorable toward j, and i will consider whether to cut j off. If λ happens to be too small, depending on L_e, the cut-off decision may be λ-robust at this very point (i.e., after the first assessment), and i will cut j off by the decision rule. Thus, λ should be properly chosen to ensure the decision process will bootstrap.

However, the look-ahead λ is related to the potential risk of being infected if the look-ahead has been carried out. Suppose that i's infection risk (against j) is $R(n)$ where n is the number of encounters between i and j; since direct contact is the only propagation channel of the proximity malware, $R(n)$ and n are positively correlated: more encounters mean a higher risk of being infected. One reasonable instantiation of $R(n)$ is $R(n) = 1-(1-p)^n$, where p is the (fixed) infection probability in a single encounter.

Suppose that i's cost of cutting j off (and hence losing j's service) is $C_i(j)$. To be comparable with the instantiation $R(n) = 1-(1-p)^n$, let $0 < C_i(j) < 1$. $C_i(j)$ reflects the value of j's service to i. One possible instantiation of $C_i(j)$ is j's social significance as perceived by i. For example, i can collect past communication/forwarding records or even initiate (opportunistic) local social community detection and use techniques such as ego-betweenness [3] to estimate j's social significance to i. The social cost $C_i(j)$ can be estimated once and kept fixed or can otherwise be updated regularly throughout the decision process.

If the evidence is unfavorable toward j, the look-ahead λ can be chosen by $\lambda = \max\{n|R(n) \leq C_i(j)\} = \max\{n|1-(1-p)^n \leq C_i(j)\}$: i is willing to give j chance (by looking λ steps ahead and hence not cutting j off immediately) as long as the infection risk (positively correlated with λ) is less than the cost of losing j's service (if j is a good neighbor). Depending on the relation between the infection risk $R(n)$ and the social cost $C_i(j)$, λ can be either static or dynamic across multiple encounters. To put it another way, a large λ is chosen as long as the (potential) benefit of maintaining connection with j justifies the (infection) risk.

References

[1] E. Jaynes, "Information theory and statistical mechanics. ii," *Phys. Rev.*, vol. 108, no. 2, pp. 171–190, 1957.

[2] C. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. Jrnl.*, vol. 28, no. 4, pp. 656–715, 1949.

[3] E. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant MANETs," *IEEE TMC*, vol. 8, no. 5, pp. 606–621, 2009.