DETECTING ENCRYPTED INTERACTIVE STEPPING-STONE CONNECTIONS

Ting He and Lang Tong

ABSTRACT

Network intruders often hide their identities by sending attacks through a chain of compromised hosts that are used as "stepping stones". The difficulty in defending against such attacks lies in detecting stepping-stone connections at the compromised hosts. In this paper, to distinguish normal from attacking connections, we consider strategies that do not depend on the content of the traffic so that they are applicable to encrypted traffic. We propose a low complexity detection algorithm that has no miss detection and an exponentially-decaying false alarm probability. A sequential strategy is then developed to reduce the required number of testing packets.

Keywords: Stepping-stone detection, intrusion detection algorithms, encrypted stepping-stone attacks, interactive stepping-stone attacks.

1. INTRODUCTION

Stepping-stone attack is a common way for network intruders to conceal their identity. In a stepping-stone attack, the attacker compromises (multiple) hosts as relay machines, constructs a chain of connections through these hosts using remote login such as Telnet or SSH, and then sends attacking commands through this chain to the victim [1]. Because each connection is made by a separate remote login, the next host in the chain can only see the identity of its immediate upper stream neighbor, and the victim only sees the identity of the last host. Therefore, we have to trace back the chain to find the origin of an attack. Such tracing can be overwhelming because of the huge volume and highly dynamic nature of the network traffic.

To address this issue, Donoho *et al.* propose in [2] to install stepping-stone monitors at each gateway node to detect stepping-stone pairs[1] by examining the incoming/outgoing traffic. In practice, the monitor has to make decisions by observing live traffic, which may not include the beginning or the end of the connection. Therefore, it is desirable that the detection strategy does not require synchronization between incoming and outgoing streams. Besides, the connections may be encrypted (*e.g.*, SSH sessions) so that the monitor cannot rely on the content of the traffic. Furthermore, a careful attacker may even actively modify the traffic each time it passes through a host in order to confuse the monitor.

1.1. Related Work

Staniford and Heberlein [1] are the first to consider the problem of stepping-stone connection detection. The early work is based on the content of the traffic, *e.g.*, see [1, 3]. These techniques help to recognize connections on the same intrusion path by analyzing the content of the packets. Later on timing characteristics of the traffic are used to detect encrypted stepping-stone connections, examples of which include [4–6]. The drawback of these techniques is that they are vulnerable to the active timing perturbation by the attacker.

There are a few results on detecting encrypted, perturbed stepping-stone connections; see [2, 7, 8]. The key assumption of these approaches is that there is a limit on the attacker's ability to alter the traffic. Specifically, in [2] it is assumed that there is a *maximum tolerable delay* for attacking packets, in [7] the attacker's timing perturbation is independent and identically distributed across packets, and in [8] there are constraints not only on the maximum delay, but also on the maximum number of packets that can be sent during the delay. From an algorithmic point of view, Blum, Song and Venkataraman [8] develop the first detection algorithms which require provable (polynomial) sample sizes to achieve certain false alarm probabilities.

1.2. Summary of Results and Organization

Our work is based on the same assumptions as in [8]. In this paper, we consider detecting encrypted, interactive stepping-stone connections. By "encrypted" we mean that we cannot use the content of packets. "Interactive connections" means that the new commands to be sent depend on the feedback of previous commands. Therefore, the attacker cannot wait too long for the packets to be relayed, and cannot issue packets too fast because he needs time to process the feedback. With these constraints, we reduce the problem to testing pairs of independent point processes against relayed point processes with bounded delay and bounded peak rate.

We take an algorithmic approach. Noticing that under the bounded delay and bounded peak rate assumption, the maximum variation (defined later) for stepping-stone pairs is always bounded, we develop a detection algorithm based on the maximum variation statistics. The algorithm has no miss detection and an exponentially-decaying false alarm probability. Moreover, we explore the possibility of reducing sample size by using sequential detection. Specifically, we propose an iterative algorithm which distributes the total false alarm constraint among iterations. For given false alarm constraint, we show how to decide the detection threshold adaptively

T. He and L. Tong ({th255,lt35}@cornell.edu) are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853.

This work is supported in part by TRUST (The Team for Research in Ubiquitous Secure Technology), which receives support from the National Science Foundation (NSF award number CCF-0424422) and the following organizations: Cisco, ESCHER, HP, IBM, Intel, Microsoft, ORNL, Qualcomm, Pirelli, Sun and Symantec, and the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

[1] A pair of incoming and outgoing streams is called a *stepping-stone pair* if it is part of a stepping-stone attack. Otherwise, the pair is normal.

so that the constraint is satisfied. We also consider how to distribute the false alarm constraint to minimize the maximum sample size, and we show that the minimax distribution reduces the sequential algorithm to a fixed-sample-size algorithm. Our analysis focuses on the error exponent of various algorithms. We show that although our algorithm needs the same order of sample size as the algorithm proposed in [8], our algorithm has a much larger error exponent.

The rest of the paper is organized as follows. Section 2 defines the problem. Section 3 presents several detection algorithms together with the performance analysis and comparison. There are also simulation results to verify our analysis.

2. THE PROBLEM STATEMENT

Let $S_1,\ S_2$ ($S_i = (\ldots,\ s_{-1}^{(i)},\ s_0^{(i)},\ s_1^{(i)}, \ldots),\ i = 1,\ 2$) be the incoming and outgoing streams at a particular gateway node[2], and $\mathcal{T}_i = \{\ldots,\ s_{-1}^{(i)},\ s_0^{(i)},\ s_1^{(i)}, \ldots\}$ be the set of the elements in S_i. Assume that if $(S_1,\ S_2)$ is a normal pair, they are independent Poisson processes. If $(S_1,\ S_2)$ is a stepping-stone pair, then there exists a bijection $g : \mathcal{T}_1 \to \mathcal{T}_2$ such that $0 \le g(s) - s \le \Delta$ for any $s \in \mathcal{T}_1$; furthermore, $|\{s \in S_1 :\ s \in [t,\ t+\Delta]\}| \le p_\Delta$ for any t. Here Δ is the maximum tolerable delay, and p_Δ is the largest number of packets the attacker can send within Δ[3]. We want to test the following binary hypotheses:

$$\mathcal{H}_0 : \quad (S_1,\ S_2) \text{ is a normal pair,} \tag{1}$$

$$\mathcal{H}_1 : \quad (S_1,\ S_2) \text{ is a stepping-stone pair.} \tag{2}$$

by observing $(s_1^{(i)},\ s_2^{(i)},\ s_3^{(i)}, \ldots)$ $(i = 1,\ 2)$.

3. DETECTION ALGORITHMS AND PERFORMANCE ANALYSIS

Merge $(s_1^{(1)},\ s_2^{(1)}, \ldots)$ and $(s_1^{(2)},\ s_2^{(2)}, \ldots)$ and order the union as $s_1, s_2, s_3, \ldots$. Let $N_i(w) (i = 1, 2)$ be the number of packets from S_i when the total number of packets is w, *i.e.,* $N_i(w) \triangleq \sum_{j=1}^{w} I_{s_j \in S_i}$, where $I_{\cdot}$ is the indicator function. Define the cumulative difference $d(w)$ and the maximum variation $v(w)$ as

$$d(w) \triangleq N_1(w) - N_2(w), \qquad v(w) \triangleq \max_{1 \le i \le w} d(i) - \min_{1 \le i \le w} d(i).$$

3.1. DETECT-MAXIMUM-VARIATION (DMV)

Given interval I, let $N_i(I)$ be the number of packets on S_i in the interval I. We notice that the stepping-stone pairs have bounded difference in $N_i(I)$, as stated in the following proposition:

Proposition 3.1 *For stepping-stone pairs, we have*

$$|N_1(I) - N_2(I)| \le p_\Delta \quad \forall \text{ interval } I.$$

Proof: Let $I = [a,\ b]$. $N_1(I)$ is the number of incoming packets in I. By the bounded delay assumption, $N_2(I)$ can include at most the packets transmitted in $[a - \Delta,\ b]$, and at least those transmitted in $[a,\ b - \Delta]$; see Fig. 1.

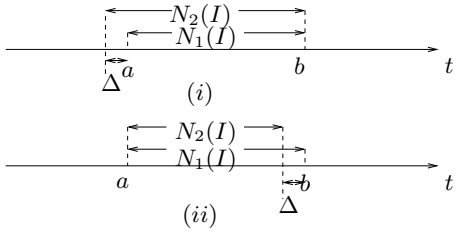

Fig. 1. (i): $N_2(I) > N_1(I)$; (ii): $N_1(I) > N_2(I)$.

Therefore, for any interval I, $-p_\Delta \le N_1(I) - N_2(I) \le p_\Delta$. ■

We can use the maximum difference $|N_1(I) - N_2(I)|$ over all I to detect stepping-stone pairs. Noticing that

$$\max_{1 \le i \le j \le w} |(N_1(j) - N_1(i)) - (N_2(j) - N_2(i))| = v(w),$$

we can equivalently use the maximum variation to detect stepping-stone pairs. The algorithm is shown in Table 3.1.

DETECT-MAXIMUM-VARIATION($S_1,\ S_2,\ p_\Delta,\ n$):

$d_{\max} = d_{\min} = 0$;
for $w = 1 : n$
  $d(w) = \begin{cases} d(w-1) + 1 & \text{if } s_w \in S_1 \\ d(w-1) - 1 & \text{if } s_w \in S_2 \end{cases}$;
  $d_{\max} = \max(d_{\max},\ d(w))$;
  $d_{\min} = \min(d_{\min},\ d(w))$;
  if $d_{\max} - d_{\min} > p_\Delta$ return NORMAL;
end
return ATTACK;

Table 1. DETECT-MAXIMUM-VARIATION (DMV).

Algorithm DMV has time complexity $O(n)$ and uses only constant memory ($O(\log p_\Delta)$, to be precise). By Proposition 3.1, any stepping-stone pair will be detected after n packets, *i.e.,* miss detection is totally avoided. We only need to be concerned about the false alarm probability, which is bounded as follows.

Theorem 3.2 *The false alarm probability of DMV is bounded by*

$$P_F(DMV) \le \frac{(p_\Delta + 1)}{1 - \rho} \rho^n,$$

where $\rho = \cos \frac{\pi}{p_\Delta + 2}$. *Furthermore,*

$$\lim_{n \to \infty} -\frac{1}{n} \log P_F(DMV) = -\log \rho.$$

Proof: See Appendix. ■

[2] $s_k^{(i)}$ is the arrival epoch of the kth packet on stream i since the monitor starts (if $k \le 0$, it is the $(-k+1)$th packet before the monitor starts).

[3] The notion of Δ and p_Δ is first used in [2] and [8] respectively. Here we do not consider inserting chaff packets. See [2] for the description of such scenario.

Remarks: For given false alarm constraint δ, making the upper bound in Theorem 3.2 equal to δ yields a sample size

$$n = \frac{\log \delta(1-\rho) - \log(p_\Delta + 1)}{\log \rho} = O\left(p_\Delta^2 \log \frac{p_\Delta}{\delta}\right).$$

Blum *et al.* [8] proposed an algorithm called "DETECT-ATTACKS" (DA) for stepping-stone detection. Algorithm DA divides samples into groups of $2(p_\Delta + 1)^2$ packets each, and computes the cumulative difference $d(w)$ for each group. It returns NORMAL if $|d(w)| > p_\Delta$ in any of the groups. Blum *et al.* prove that $2(p_\Delta + 1)^2 \log \frac{1}{\delta}$ packets are required to guarantee a false alarm probability no more than δ.

We point out that DMV always outperforms DA[4]. The reason is that since $v(w) \geq \max_{1 \leq i \leq w} |d(i)|$, for every realization, if DMV has a false alarm, DA must have a false alarm too.

Now we compare their false alarm probabilities. We have the following lemma:

Lemma 3.3 *For independent Poisson processes and large p_Δ,*

$$P_F(DA) \geq \left(K^{\frac{1}{2(p_\Delta+1)^2}} \sigma\right)^n \tag{3}$$

where $\sigma = \cos \frac{\pi}{2(p_\Delta+1)}$, and $K = \frac{\sin \frac{\pi}{2(p_\Delta+1)}}{2(p_\Delta+1)(1-\sigma)}$.

Proof: See Appendix. ∎

From Lemma 3.3 we have that for large p_Δ, the error exponent of DA is at most $-\log(K^{\frac{1}{2(p_\Delta+1)^2}} \sigma)$. By Taylor expansion,

$$\begin{aligned} -\log \rho &= \frac{\pi^2}{2(p_\Delta+2)^2} + o\left(\frac{1}{p_\Delta^2}\right), \\ -\log\left(K^{\frac{1}{2(p_\Delta+1)^2}} \sigma\right) &= \frac{\frac{\pi^2}{4} + \log \frac{\pi}{2}}{2(p_\Delta+1)^2} + o\left(\frac{1}{p_\Delta^2}\right). \end{aligned}$$

Therefore, for large p_Δ, the false-alarm error exponent of DMV is at least 3.38 times larger than that of DA.

Fig. 2 plots the simulated false alarm probabilities of DA and DMV and their bounds[5]. It confirms our claim that the false alarm probability of DMV decays much faster than that of DA.

3.2. SEQUENTIAL-DMV (SDMV)

In both DA and DMV, the decision of ATTACK requires a fixed sample size. We hope to make ATTACK decisions sequentially so that we can possibly use fewer packets. To this end, we propose to use an iterative algorithm and divide the total false alarm constraint among iterations. Specifically, we split the total false alarm probability δ into $q_1\delta,\ q_2\delta, q_3\delta, \ldots$, where $\mathbf{q} = (q_1,\ q_2, \ldots)$ satisfies $q_w \geq 0$ and $\sum_{w=1}^{\infty} q_w = 1$. If, in each iteration w, the false alarm probability is bounded by $q_w\delta$, then by union bound we see that the total false alarm will be bounded by δ.

[4]Note that in terms of the order of sample size with respect to p_Δ and δ, DMV and DA are comparable.

[5]Note that the sample size of DA has to be a multiple of the group size $2(p_\Delta + 1)^2$.

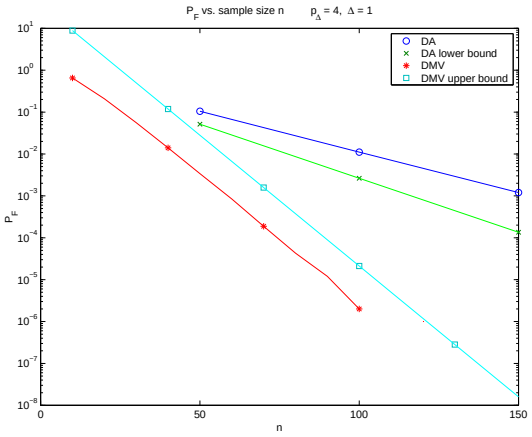

Fig. 2. P_F of algorithms DA and DMV. (Assume normal pairs are independent Poisson processes with the same rate.)

Using $v(w)$ as statistics, we obtain from the proof of Theorem 3.2 that, for normal pairs,

$$\Pr\{v(w) < \tau_w\} \leq \frac{\tau_w \left(\cos \frac{\pi}{\tau_w+1}\right)^w}{1 - \cos \frac{\pi}{\tau_w+1}} \triangleq f(\tau_w, w), \text{ for any } \tau_w \geq 1.$$

If $\tau_w \triangleq \sup\{\text{integer } k:\ f(k,\ w) \leq q_w\delta\}$, and the decision rule is to return ATTACK if $v(w) < \tau_w$, then the false alarm probability of the wth iteration is bounded by $q_w\delta$. Therefore we have the sequential algorithm in Table 3.2.

SEQUENTIAL-DMV($S_1,\ S_2,\ p_\Delta,\ \delta,\ \mathbf{q}$):

$d_{\max} = d_{\min} = 0$;
for $w = 1,\ 2, \ldots$
$d(w) = \begin{cases} d(w-1) + 1 & \text{if } s_w \in S_1 \\ d(w-1) - 1 & \text{if } s_w \in S_2 \end{cases}$;
$d_{\max} = \max(d_{\max},\ d(w))$;
$d_{\min} = \min(d_{\min},\ d(w))$;
if $d_{\max} - d_{\min} > p_\Delta$ return NORMAL;
$\tau = \sup\{\text{integer } k:\ f(k,\ w) \leq q_w\delta\}$;
if $d_{\max} - d_{\min} < \tau$ return ATTACK;
end

Table 2. SEQUENTIAL-DMV (SDMV).

Algorithm SDMV also uses the maximum variation as the statistic, and therefore can be thought of as a sequential version of DMV. The vector $\mathbf{q}$ is part of the algorithm design. Ideally, we want to choose $\mathbf{q}$ to minimize the average sample size. If the attacker's strategy is unknown, then we may wish to minimize the largest sample size. Specifically, if the attacker does the best to evade detection by keeping $v(w) = p_\Delta$ for all $w \geq p_\Delta$, then the best $\mathbf{q}$ is

$$q_n = 1 \quad \text{for } n = \inf\left\{w:\ \frac{(p_\Delta+1)\left(\cos \frac{\pi}{p_\Delta+2}\right)^w}{1 - \cos \frac{\pi}{p_\Delta+2}} \leq \delta\right\}$$

and $q_w = 0$ for all $w \neq n$. That is, the minimax $\mathbf{q}$ reduces SDMV to the fixed-sample-size algorithm DMV.

4. CONCLUSION

In this paper, we consider detecting stepping-stone connections with bounded delay and bounded peak rate. Our techniques can rule out independent connection pairs with provable confidence and hopefully leave a much smaller number of suspicious connections for further examination. Therefore, they are most useful in the scenario when the total volume of the traffic to be analyzed is large. In [9], we consider detecting stepping-stone connections with bounded delay only or bounded memory. These more general assumptions will make the detection techniques easier to apply in practice.

5. APPENDIX

5.1. Proof of Theorem 3.2 and Lemma 3.3

Proof:

The proof is based on the theory of random walk. Let $\{X_n\}_{n\geq 0}$ be a simple random walk, *i.e.*,

$$X_0 = 0, \quad X_n = Z_1 + Z_2 + \ldots + Z_n, \ (n > 0)$$

where $\{Z_i\}_{i=1,\,2,\ldots}$ are i.i.d. random variables taking value in $\{-1,\ 0,\ 1\}$. Let $p = \Pr\{Z_i = 1\}$, $q = \Pr\{Z_i = -1\}$. Define *the hitting time* of $-b$ or a ($a,\ b \geq 0$) as

$$N_{-b,\,a} = \inf\{n \geq 1 :\ X_n = -b \text{ or } a\}.$$

In [10], it is proved that

$$\Pr\{N_{-b,\,a} = n\} \leq \frac{1}{2}\left(\frac{p}{q}\right)^{a/2} \frac{1}{s_1^{n-1}} + \frac{1}{2}\left(\frac{q}{p}\right)^{b/2} \frac{1}{s_1^{n-1}}, \quad (4)$$

where $s_1 = \frac{1}{1-p-q+2(pq)^{\frac{1}{2}}\cos\left(\frac{\pi}{a+b}\right)}$, and

$$\lim_{n\to\infty} -\frac{1}{n}\log \Pr\{N_{-b,\,a} > n\} = \log s_1. \quad (5)$$

If $a = b$, then for large n,

$$\Pr\{N_{-b,\,a} = n\} \geq \frac{\sin\frac{\pi}{2a}}{2a s_1^{n-1}}. \quad (6)$$

For the proof of Theorem 3.2, note that for independent Poisson processes, it is known that $d(w)$ is a simple random walk. Define extreme values $U_n = \max_{i=0,\ldots,\,n} d(i)$, $L_n = \min_{i=0,\ldots,\,n} d(i)$. A false alarm occurs in DMV if and only if $U_n - L_n < p_\Delta + 1$. Note that the false alarm probability is the largest if $d(w)$ is symmetric (*i.e.*, $p = q = \frac{1}{2}$). We have

$$\begin{aligned} P_F(\text{DMV}) &= \Pr\{U_n - L_n < p_\Delta + 1\} \\ &= \Pr\{\bigcup_{a=1}^{p_\Delta+1} \{U_n < a,\ L_n > -(p_\Delta + 2 - a)\}\} \\ &\leq \sum_{a=1}^{p_\Delta+1} \Pr\{U_n < a,\ L_n > -(p_\Delta + 2 - a)\} \quad (7) \\ &\leq (p_\Delta + 1)\frac{\rho^n}{1-\rho}, \quad (8) \end{aligned}$$

where $\rho = \cos\frac{\pi}{p_\Delta+2}$. Here (7) is by union bound, and (8) is by noticing

$$\Pr\{U_n < a,\, L_n > -(p_\Delta + 2 - a)\} = \Pr\{N_{-(p_\Delta+2-a),\,a} > n\},$$

and then applying (4) with $p = q = \frac{1}{2}$. Furthermore, by (5) it is easy to see that $\lim_{n\to\infty} -\frac{1}{n}\log P_F(\text{ DMV}) = -\log\rho$.

For the proof of Lemma 3.3, note that DA has false alarm in a given group if and only if the maximum $|d(i)|$ in that group is within p_Δ. Observing that

$$\Pr\{\max_{i\in\{1,\ldots,\,2(p_\Delta+1)^2\}} |d(i)| \leq p_\Delta\} = \Pr\{N_{-(p_\Delta+1),\,(p_\Delta+1)} > 2(p_\Delta+1)^2\},$$

we apply (6) with $a = b = p_\Delta + 1$ to lower bound the false alarm of DA on one group. Then taking the product over all groups gives the desired result.

■

6. REFERENCES

[1] S. Staniford-Chen and L. Heberlein, "Holding intruders accountable on the internet," in *Proc. the 1995 IEEE Symposium on Security and Privacy*, (Oakland, CA), pp. 39–49, May 1995.

[2] D. Donoho, A. Flesia, U. Shankar, V. Paxson, J. Coit, and S. Staniford, "Multiscale stepping-stone detection: Detecting pairs of jittered interactive streams by exploiting maximum tolerable delay," in *5th International Symposium on Recent Advances in Intrusion Detection, Lecture Notes in Computer Science 2516*, 2002.

[3] X. Wang, D. Reeves, S. Wu, and J. Yuill, "Sleepy watermark tracing: An active network-based intrusion response framework," in *Proc. of the 16th International Information Security Conference*, pp. 369–384, 2001.

[4] Y. Zhang and V. Paxson, "Detecting stepping stones," in *Proc. the 9th USENIX Security Symposium*, pp. 171–184, August 2000.

[5] K. Yoda and H. Etoh, "Finding a connection chain for tracing intruders," in *6th European Symposium on Research in Computer Security, Lecture Notes in Computer Science 1895*, (Toulouse, France), October 2000.

[6] X. Wang, D. Reeves, and S. Wu, "Inter-packet delay-based correlation for tracing encrypted connections through stepping stones," in *7th European Symposium on Research in Computer Security, Lecture Notes in Computer Science 2502*, pp. 244–263, 2002.

[7] X. Wang and D. Reeves, "Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays," in *Proc. of the 2003 ACM Conference on Computer and Communications Security*, pp. 20–29, 2003.

[8] A. Blum, D. Song, and S. Venkataraman, "Detection of Interactive Stepping Stones: Algorithms and Confidence Bounds," in *Conference of Recent Advance in Intrusion Detection (RAID)*, (Sophia Antipolis, French Riviera, France), September 2004.

[9] T. He and L. Tong, "Detecting Encrypted Stepping-stone Connections," Tech. Rep. ACSP-TR-01-06-02, Cornell University, January 2006. `http://acsp.ece.cornell.edu/pubR.html`.

[10] D. Cox and H. Miller, *The Theory of Stochastic Processes*. New York: John Wiley & Sons Inc., 1965.