Stochastic Models for Remote Timing Attacks

Simone Bozzolan, Diletta Olliaro, Stefano Calzavara, Andrea Marin
Università Ca' Foscari Venezia, Venice, Italy

Gianfranco Balbo, Matteo Sereno
Università di Torino, Turin, Italy

***Abstract*—We introduce the first model-based remote timing attack that combines queueing theory and Bayesian classification to infer service times of different classes of network requests. Unlike empirical methods, our approach calculates the posterior probability that an observed service time belongs to a target class, enabling precise attack decisions with quantified confidence. Our experiments on popular web applications and websites show that our investigation is not just a theoretical exercise, because our attack outperforms existing empirical approaches in terms of standard performance figures.**

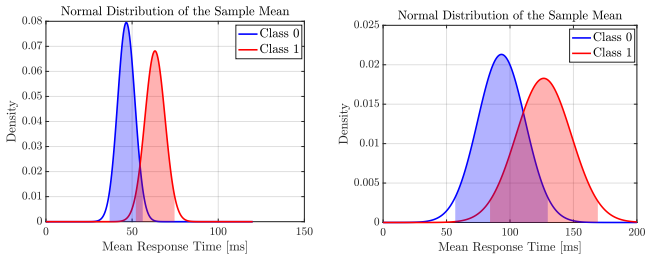
I. Introduction

Remote timing attacks exploit response time variations to infer private information [1], such as user tracking [2] and authentication state [3]. Previous work relies on empirical methods, lacking theoretical grounding. We propose a stochastic modeling approach using queueing theory [4], [5] and Bayesian inference to classify network requests from response times, improving robustness by considering load variability. We will focus on cross-site timing attacks, where attackers use a victim's browser to measure response times and infer authentication state. Attacker reliance on response times faces limitations due to service time variability under different system loads. Experimental results (Fig. 1) from a WordPress local setup demonstrate authentication-state inference feasibility under low load ($\rho = 0.4$), but distribution overlap at high load ($\rho = 0.7$) makes the attack unreliable. To address this, we estimate arrival (λ) and service rates (μ) from observed data to derive service time distributions and a Bayesian classifier uses these distributions for request classification.

II. Contribution

Attack Description. Our attack operates in two phases: *exploration* and *exploitation*. During the first one, the attacker collects timing information about the target web application at different times and for different classes of requests. The exploitation phase leverages the information collected in the exploration phase to build a stochastic model of the target web application and actually mount the attack. In particular, the attacker measures the response times of requests of an unknown class and uses the model to reconstruct their class.

Queueing Model. The system is modeled using a M/M/1/PS queueing model [6], leveraging the LST of the waiting time distribution conditioned on a job's deterministic service time [7, Eq. (30)]. The latter expression is used to estimate the probability density of an observation of the waiting time given an instance of the service time (τ) and it depends on the system parameters λ (arrival rate), μ (service rate). This leads to the following steps: (i) discretizing empirical service time distributions for both classes (ii) estimation of the system parameters through observed response times, as direct measurement of system load is unavailable.



(a) $\mu = 40$ j/s, $\rho = 0.4$ (b) $\mu = 40$ j/s, $\rho = 0.7$

Fig. 1: Response time distributions with different ρ

TABLE I: Comparison with BakingTimer

Measure	BakingTimer	Our Attack
TPR	0.78	0.83
TNR	0.92	0.93
AR	0.37	0.06

Statistical Model. Finally, we leverage Bayesian statistics to answer the question: *what is the likely class of a request given an estimate of its service time?* During our attack, we collect a set of observations X^A, and we apply Bayesian estimation (Eq. (1)) to compute the probability that $\theta = \tau_\ell$, where τ_ℓ is the service time parameter of the $\ell-$th class. Let P_θ be the prior probability of parameter θ, assumed to be equiprobable for all classes if no prior knowledge is available. According to Bayesian statistics, the posterior probability of $\theta = \tau_\ell$, is:

$$g_{X^A}(\theta|x_1^A,\ldots,x_{N^A}^A) = \frac{f_X(x_1^A,\ldots,x_{N^A}^A|\theta)P_\theta}{G}, \quad (1)$$

where G is a probability normalizing constant and $f_X(x_1^A,\ldots,x_{N^A}^A|\theta)$ is the likelihood function of the sequence.

Experimental Evaluation. We test our attack on 20 popular websites from the Tranco ranking [8], comparing performance with the state-of-the-art attack BakingTimer [3] on the true positive rate (TPR), true negative rate (TNR), and abstention rate (AR). The attack targets login state inference via cross-site timing attacks. Tests are conducted at varying times to account for server load, with averaged results presented in Table I.

III. Conclusion

We propose using stochastic models to enhance remote timing attacks. Starting with a queueing model of the target application, Bayesian statistics is employed to infer the class of specific network requests. Compared to prior empirical methods, our approach is more precise, provides quantitative probability bounds, and requires fewer requests.

REFERENCES

[1] S. A. Crosby, D. S. Wallach, and R. H. Riedi, "Opportunities and limits of remote timing attacks," *ACM Trans. Inf. Syst. Secur.*, vol. 12, no. 3, pp. 17:1–17:29, 2009.

[2] E. W. Felten and M. A. Schneider, "Timing attacks on web privacy," in *CCS 2000, Proceedings of the 7th ACM Conference on Computer and Communications Security, Athens, Greece, November 1-4, 2000*, D. Gritzalis, S. Jajodia, and P. Samarati, Eds. ACM, 2000, pp. 25–32.

[3] I. Sánchez-Rola, D. Balzarotti, and I. Santos, "Bakingtimer: privacy analysis of server-side request processing time," in *Proceedings of the 35th Annual Computer Security Applications Conference, ACSAC 2019, San Juan, PR, USA, December 09-13, 2019*, D. M. Balenson, Ed. ACM, 2019, pp. 478–488.

[4] W. Wang, G. Casale, A. Kattepur, and M. Nambiar, "QMLE: A methodology for statistical inference of service demands from queueing data," *ACM Trans. Model. Perform. Evaluation Comput. Syst.*, vol. 3, no. 4, pp. 17:1–17:28, 2018.

[5] F. Baccelli, B. Kauffmann, and D. Veitch, "Inverse problems in queueing theory and internet probing," *Queueing Syst. Theory Appl.*, vol. 63, no. 1-4, pp. 59–107, 2009.

[6] L. Kleinrock, *Queueing Systems, Volume 1: Theory*. Wiley-Interscience, 1975.

[7] E. G. C. Jr., R. R. Muntz, and H. F. Trotter, "Waiting time distributions for processor-sharing systems," *J. ACM*, vol. 17, no. 1, pp. 123–130, 1970.

[8] V. L. Pochat, T. van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February 24-27, 2019*. The Internet Society, 2019, pp. 1–15.