A Novel Network Delay Based Side-Channel Attack: Modeling and Defense

Zhen Ling*, Junzhou Luo*, Yang Zhang*, Ming Yang*, Xinwen Fu†and Wei Yu‡
*Southeast University, Nanjing 211189, P. R. China. {zhenling, jluo, yangzhang, yangming2002}@seu.edu.cn
†University of Massachusetts Lowell, Lowell, MA 01854, USA. xinwenfu@cs.uml.edu
‡Towson University, Towson, MD 21252, USA. wyu@towson.edu

***Abstract*—Information leakage via side channels has become a primary security threat to encrypted web traffic. Existing side channel attacks and corresponding countermeasures focus primarily on packet length, packet timing, web object size and web flow size. However, we found that encrypted web traffic can also leak information via network delay between a user and the web sites that she visits. Motivated by this observation, we investigate a novel network-delay based side-channel attack to infer web sites visited by a user. The adversary can utilize pattern recognition techniques to differentiate web sites by measuring sample mean and sample variance of the round-trip time (RTT) between a victim user and web sites. We theoretically analyzed the damage caused by such an adversary and derived closed-form formulae for detection rate, the probability that the adversary correctly recognizes a web site. To defeat this side-channel attack, we proposed several countermeasures. The basic idea is to shape traffic from different web sites so that they have similar RTT statistics. We proposed the strategies based on the k-$means$ clustering and $\mathbb{K}$-Anonymity to ensure that traffic shaping will not cause excessive delay while providing a predictable degree of anonymity. We conducted extensive experiments and our empirical results match our theory very well.**

***Index Terms*—Network Delay, Side Channel, Information Leak, Countermeasures.**

I. Introduction

With growing concerns over privacy, security and censorship, various privacy enhancing technologies (PET) have been proposed and deployed to improve users' Internet experience [1]. Encrypted tunnels is one common strategy to hide the content and IP addresses of web sites from third-party observers. Single-hop systems, including OpenSSH tunnels, simple SSL proxies and virtual private networks (VPN) are examples of encrypted tunnels. NAT enabled wireless routers with WPA/WPA2 encryption also provide anonymity and privacy for users behind these routers to some extent. However, these technologies are insufficient to prevent a third party from taking advantage of side channels to infer critical information from encrypted web traffic and identifying web sites being visited by users.

To identify web sites visited by users, an adversary can profile the "fingerprints" of these web sites. Side channels provide such a fingerprint, and are often based on statistics of web traffic. Existing side-channel attacks against webs can be classified into two categories: object based and packet based techniques. In web object based side channels [2], an adversary explores the HTTP response and estimates the size of HTML objects, and then measures the similarity between the estimated feature and profile database. In packet based side channels [3], [4], [5], [6], [7], the adversary takes advantage of features of encrypted packets, including the packet size distribution, flow direction, packet inter-arrival times, and packet ordering. To prevent information leakage from these side channels, researchers have proposed two categories of countermeasures: server-side solution [8] and client-side solution [9]. A server-side solution morphs traffic from different web sites so that they look similar to each other. The client-side countermeasure against side channels manipulates TCP and HTTP features of web traffic.

In this paper, we investigate a novel network delay based side-channel attack against encrypted web traffic. Through the study of web traffic via the SSH tunnel, we found that network delay, i.e., the round-trip time (RTT), could leak critical information about the encrypted web traffic. Motivated by this observation, we adopt information theory to estimate the information leakage from RTTs, and establish a generic traffic classification model to identify web sites accessed by users. Sample mean and sample variance are used as statistical features of RTTs to differentiate web sites. Detection rate is defined as the probability that an adversary can correctly recognize a web site. We derive closed-form formulae for the detection rate in terms of parameters, including the mean and variance of the network delay and the sample size.

To counter the attack, we proposed three defense strategies: (i) In the statistical distribution based approach, we manipulate the web traffic delay at the proxy side so that sample mean and sample variance of the RTT for all web sites are similar. However, this strategy introduces significant delay into web traffic. (ii) In the k-$means$ clustering based strategy, we classify web sites into several clusters and manipulate web traffic from each cluster of web sites so that their sample mean and sample variance are similar. However, the k-$means$ clustering cannot predict the number of web sites in each cluster and cannot guarantee the degree of anonymity for these clusters of web sites. (iii) To guarantee the degree of anonymity for each cluster of web sites, we propose the $\mathbb{K}$-$Anonymity$ based approach. An efficient heuristic algorithm is proposed to adjust the number of web sites in each cluster to achieve $\mathbb{K}$-$Anonymity$, i.e., at least $\mathbb{K}$ web sites in each cluster. This $\mathbb{K}$-$Anonymity$ strategy provides the tradeoff between privacy and web surfing experience.

We conducted extensive real-world experiments to validate

978-1-4673-0775-8/12/$31.00 ©2012 IEEE

the severity of the novel side-channel attack and the feasibility and effectiveness of our countermeasures. Our experiments demonstrate that the $\mathbb{K}$-$Anonymity$ based strategy can effectively balance the tradeoff between the performance and information leakage. Our data match our theory very well.

The remainder of this paper is organized as follows: We review the related work in Section II, and present the network delay based side-channel attack, including network and threat models, and information leakage metric in Section III. In Section IV, we analyze the detection rate and derive the closed formulae for detection rate when the adversary uses sample mean and sample variance as the statistical features. We propose the countermeasures in Section V. Extensive experimental results are presented in Section VI and we conclude this paper in Section VII.

II. Related Work

Existing traffic analysis techniques can largely be categorized into two groups: side channels (**passively** observing traffic) and covert channels (**actively** manipulating traffic). In a side-channel attack, an attacker records traffic passively and identifies the similarity between server's outbound traffic and client's inbound traffic to correlate the communication relationship. For example, Zhu *et al*. [10] used mutual information for the similarity measurement. Levine *et al*. [11] used a cross correlation technique. Research also showed that the attacker could infer sensitive information from encrypted network traffic by examining patterns of packet size and timing. For example, Song *et al.* [12] utilized the packet inter-arrival timing in SSHv1 connections to infer keystroke patterns and ultimately crack the typed passwords. Sun *et al.* [13] investigated the sizes of HTML objects transmitted over SSL connections and were able to identify the web pages based on the number and size of objects in each encrypted HTTP response. Liberatore and Levine [4] examined packet size of HTTP traffic transmitted over the persistent connection or tunneled via SSH that could statistically identify the web pages. Lu *et al.* [6] improved the web site fingerprinting techniques by utilizing packet size ordering information. Wright *et al.* [14] investigated the packet size statistical distribution in encrypted Voice over IP (VoIP) connections and could identify the spoken language. Later work of Wright *et al.* [15] also investigated how an eavesdropper could identify spoken phrases in encrypted VoIP.

The covert channel techniques intend to embed a specific secret signal into target traffic. Such techniques can reduce the false positive rate significantly if the signal is long enough and does not require massive traffic training as required in side-channel techniques. For example, Wang *et al.* [16] proposed an active watermarking scheme that was robust to random timing perturbation. They analyzed the tradeoffs between the true positive rate, the maximum timing perturbation added by attackers, and the number of packets needed to successfully decode the watermark. Wang *et al.* [17] investigated the feasibility of a timing-based watermarking scheme to identify the encrypted peer-to-peer VoIP calls. By slightly changing the timing of packets, their approach can correlate encrypted anonymous network connections. Peng *et al.* [18] analyzed the secrecy of timing-based watermarking techniques proposed in [16], based on the distribution of traffic timing. Yu *et al.* [19] proposed a flow marking scheme based on the direct sequence spread spectrum (DSSS) technique. This technique could be used by attackers to secretly confirm the communication relationship via mix networks. Ling *et al.* [20] proposed the cell counter based attack against Tor, in which an attacker embeds a signal into the variation of cell counter of the target traffic at the malicious exit onion router. Ling *et al.* [21] enhanced the approach without controlling the entry and exit node. Ling *et al.* [22] investigated the least significant packet size based attack against Anonymizer.

To counter traffic analysis attacks, Wright *et al.* [8] proposed a server-side based traffic morphing approach in which they modulate the source packet size so that packet size distributions in different traffic flows are similar to each other [14], [4]. However, this technique is vulnerable to the enhanced web fingerprinting techniques [6]. Luo *et al.* [9] investigated a client-side based composite traffic transformation techniques by exploiting the basic protocol features of TCP and HTTP to manipulate the characteristics of encrypted web traffic. Unfortunately, the authors did not consider the side-channel technique caused by the RTTs between clients and web sites. This paper addresses the RTT based side channel.

III. A Network Delay Based Side-Channel Attack

In this section, we first present network and threat models. We then introduce the novel side-channel attack via network delay against encrypted web traffic for interring the corresponding web sites.

A. Network and Threat Models

Figure 1 gives our network model and threat model. It has four participants: a victim client, an attacker, a remote proxy and web sites. The client connects to a remote proxy over an encrypted transport layer. The proxy makes requests on behalf of the client, and returns the results over the encrypted connection to the client. We assume that the adversary is capable of recording the timestamp and length of packets. In addition, we assume that the proxy does not intentionally introduce additional delay and perturb the passing traffic. The proxy that we use to exemplify the side-channel attack is a SOCKS based OpenSSH implementation. However, network delay based side-channel attacks can be applied against VPN proxies and WEP/WPA wireless routers as well.

In order to measure RTT between a victim user and a remote web site, the adversary observes the communication and utilizes appropriate packets to derive RTT. After a client is authenticated by the proxy, a SSH tunnel is built between the client and proxy. When the web browser accesses the web site through the local SOCKS proxy, the client sends a "SSH_MSG_CHANNEL_OPEN" packet to the SSH proxy. The proxy receives the packet and initiates a TCP connection to the web site on behalf of the client. Once the connection is established, the proxy sends back

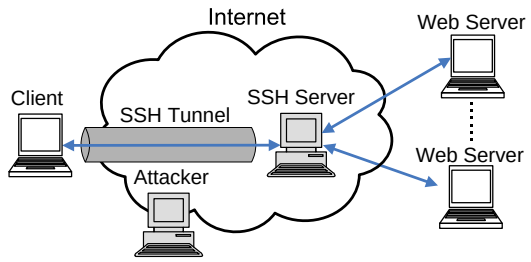

Fig. 1. Network Model and Threat Model

a "SSH_MSG_CHANNEL_OPEN_CONFIRMATION" packet to the client. Subsequently, the client transmits an HTTP "GET" packet to retrieve the web object and the proxy forwards the packet to the web site, which sends back the HTTP response packet.

This basic workflow is illustrated in Figure 2. We assume client C accesses the remote web site W via proxy P. Denote the RTT between the client C and proxy P as T_{CP}, and the RTT between the proxy P and web site W as T_{PW}. Assume the packet processing time at proxy is T_P. Then the RTT between client C and web site W can be written as follows,

$$T_{CW} = T_{CP} + T_P + T_{PW}. \quad (1)$$

Our threat model is also illustrated in Figure 2. To derive T_{CW}, the adversary can measure the time lapse between the HTTP "GET" and the corresponding response. In this way, the adversary is capable of building RTT profiles between a client and a list of known web sites. By using these profiles, the adversary attempts to decide which of the known RTTs most closely matches the encrypted and unknown RTT. The technique will be further explained in details in Section III-C.

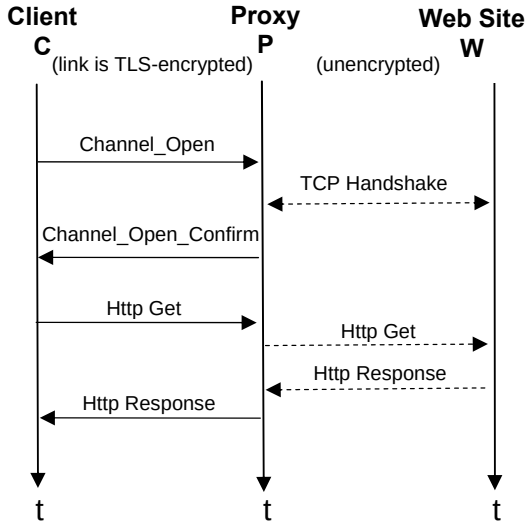

Fig. 2. Workflow of SSH Connection Establishment

B. Information Leakage Metrics

We adopt the information theoretic metrics to estimate the information leakage via the RTT based side channel. To define *anonymity*, we adopt the definition by *Pfitzmann* and *Köhntopp* in [23]. Anonymity is the state of being not identifiable within a set of subjects, the anonymity set. A sender is identifiable when we get information that can be linked to him. In our context, the "sender" is the web site that a client is accessing. Hence, we only consider the sender's anonymity in this paper. We consider an adversary who has obtained the profiles of different websites a client may visit. *Anonymity degree* is the probability that the adversary recognizes the correct web site that the client is visiting. Denote the potential web sites that the client accesses as $\{W_1, W_2, \ldots, W_N\}$, where N is the total number of web mi

2) The kernel estimator of a density function with kernel G is defined by

$$f(s) = \frac{1}{Mh}\sum_{i=1}^{M} G(\frac{s - S_i}{h}), \quad (4)$$

where h is the window width, also called the smoothing parameter or bandwidth, S_i is the i^{th} measurement of the feature, M is the number of such measurements, and $G(.)$ is based on Gaussian kernel, i.e., $G(s) = \frac{1}{\sqrt{2\pi}} exp(-\frac{s^2}{2})$.

3) Based on the PDFs of statistical features for different web sites, Bayes decision rules are derived. Recall that there are N possible web sites $W_1, \ldots, W_N$. The Bayes decision rule can be stated as follows:
The sample is classified as one from web site W_i if

$$P(W_i|s) \geqslant P(W_j|s). \quad (5)$$

That is,

$$f(s|W_i)P(W_i) \geqslant f(s|W_j)P(W_j), \quad (6)$$

for all $j = 1, \ldots, N$. Here $P(W_i)$ is a priori probability that the web site W_i can be accessed, and $P(W_i|s)$ is the post priori probability that the web site W_i is accessed by the user when the collected sample has the measured feature s.

Phase 2: Run-time Classification Once the profiling phase is complete, the adversary is ready to start the classification at run time. The method in Section III-A can be used to collect a sample of RTT between the client and remote web sites. Then the adversary calculates the value of the statistical feature from the collected sample, and utilizes the Bayes decision rules derived in the profiling phase to match the collected sample to one of the known web sites.

IV. Attack Analysis

In this section, we analyze the effectiveness of the network delay based side-channel attack in terms of detection rate. We first formally define the detection rate as the probability that the adversary can correctly classify the target traffic, i.e., identify the web site. We then derive the closed-form formulae of detection rate when the adversary uses sample mean and sample variance as the statistical feature, respectively. Our formulas will be approximate ones due to the complexity of the problem. However, these formulas correctly show the impact of the side-channel attack, in terms of different parameters including sample size and statistical feature used. In addition, the theoretical analysis plays a critical important role in the design of countermeasures in Section V.

A. Bayes Decision Rule

Without loss of generality, we analyze the classification of two web sites, W_i and W_j. Figure 3 shows the PDFs of the statistical features conditioned on the two web sites. Let d be the solution of the following equation,

$$f(W_i|s) = f(W_j|s). \quad (7)$$

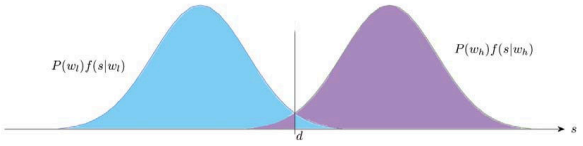

Fig. 3. Bayes Decision Making for the Case of Two Web Sites

Assume that there is a unique solution to the equation. The Bayes decision rule now becomes

$$If\ s \leqslant d,\ the\ web\ site\ is\ W_i;$$
$$otherwise,\ the\ web\ site\ is\ W_j. \quad (8)$$

The error rate P_e can then be calculated by

$$P_e = P(W_i)\int_{-\infty}^{d} f(s|W_i)ds + P(W_j)\int_{d}^{+\infty} f(s|W_j)ds. \quad (9)$$

The detection rate is given by

$$P_D = 1 - P_e = P(W_i)\int_{d}^{+\infty} f(s|W_i)ds + P(W_j)\int_{-\infty}^{d} f(s|W_j)ds. \quad (10)$$

We use three distinct features, that is, sample mean, sample variance, and the vector of sample mean and sample variance for the classification. By using these features, the adversary carries out the attack and our experimental results in Section VI demonstrate that the RTT based side-channel attack is a severe threat to user privacy while existing defense techniques cannot address this challenge.

B. Decomposition of the RTT

We formally analyze the composition of the RTT between a user and the web sites. Recall that the adversary collects a sample of RTT at run time in order to perform the classification. To derive detection rate, we need to formally model RTT. Let random variable T be the RTT between the client and remote web sites. Let T_c be the RTT between the client and the proxy and T_{w_x} be the RTT between the proxy and remote web site W_x. Let T_p be the queueing time of the proxy. Then the RTT T can be derived by

$$T = T_c + T_p + T_{w_x}. \quad (11)$$

We assume that T_c, T_p and T_{w_x} are normally distributed. Notice that this assumption simplifies our analysis without loss of generality and will be validated by our experiments in Section VI. Therefore,

$$T_c \sim \mathcal{N}(\mu_c, \sigma_c^2), \quad (12)$$
$$T_p \sim \mathcal{N}(\mu_p, \sigma_p^2), \quad (13)$$
$$T_{w_x} \sim \mathcal{N}(\mu_{w_x}, \sigma_{w_x}^2). \quad (14)$$

Denote T_i and T_j as RTT random variables when a client visits web sites W_i and W_j, respectively. Denote μ_i and μ_j as the mean of T_i and T_j. σ_i and σ_j are the variance of T_i and T_j. We have

$$T_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad (15)$$

where $\mu_i = \mu_c + \mu_p + \mu_{w_i}$ and

$$\sigma_i^2 = \sigma_c^2 + \sigma_p^2 + \sigma_{w_i}^2. \tag{16}$$

$$T_j \sim \mathcal{N}(\mu_j, \sigma_j^2), \tag{17}$$

where $\mu_j = \mu_c + \mu_p + \mu_{w_j}$ and

$$\sigma_j^2 = \sigma_c^2 + \sigma_p^2 + \sigma_{w_j}^2. \tag{18}$$

We also introduce the ratio

$$r = \frac{\sigma_i^2}{\sigma_j^2} = \frac{\sigma_c^2 + \sigma_p^2 + \sigma_{w_i}^2}{\sigma_c^2 + \sigma_p^2 + \sigma_{w_j}^2}, \tag{19}$$

where σ_c, σ_p, σ_{w_i} and σ_{w_j} are defined in Equations (12), (13), (16), (18), respectively.

C. Detection Rate

Based on the results in Section VI, we conclude that sample mean and sample variance of the RTTs can be used to identify the web sites accessed by users effectively. We now derive their detection rate formulas.

1) The Case of Sample Mean: Denote $\{X_1, X_2, \ldots, X_n\}$ as a sample of RTT. The sample mean is the average of the elements in the sample, then we can get

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}. \tag{20}$$

Notice that sample mean $\bar{X}$ is a random variable, and an unbiased estimation of X's mean μ. Then we derive a closed-form formula for the estimation of detection rate when the adversary uses sample mean as the feature statistic, which is stated in Theorem 1. The detailed proof is in Appendix A.

Theorem 1. *When sample mean is used as the feature statistic, the detection rate can be estimated by*

$$P_d(\bar{X}) \approx e^{-1/4(\mu_j - \mu_i)^2 \frac{n}{\sigma_i^2 + \sigma_j^2}} \frac{1}{\sqrt{2(\frac{1}{\sqrt{r}} + \sqrt{r})}}, \tag{21}$$

where r is defined in Equation (19)*.*

From Theorem 1 the following observations can be made:

- Detection rate in Equation (21) is increasing with sample size n. That is, if an adversary can collect a large enough sample of RTT from the same web site, the web site will be disclosed at last if different web sites have different means of RTT.
- Detection rate $P_d(\bar{X})$ is an increasing function of the difference of means $\mu_j - \mu_i$ of two web sites. If two web sites have different RTT means, the adversary can use sample mean to easily differentiate them. Later our experiments will demonstrate the significance of this observation.

2) The Case of Sample Variance: Denote $\{X_1, X_2, \ldots, X_n\}$ as a random sample of RTT of size n from the distribution X. The sample variance Y is in (22)

$$Y = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}. \tag{22}$$

Notice sample variance Y is a random variable, and an unbiased estimation of X's variance. Then we derive a closed-form formula based on *Chebyshev inequality* for the estimation of detection rate when sample variance is used as the feature statistic. The estimation of detection rate can be derived by the following theorem. The detailed proof of Theorem 2 can be found in Appendix B of our technical report [24].

Theorem 2. *Using sample variance of sample size n as the classification feature gives rise to an estimated detection rate $P_d(Y)$ as*

$$P_d(Y) \approx \max(1 - \frac{V_Y}{n-1}, 50\%), \tag{23}$$

where V_Y is calculated in Equation (24)*.*

$$V_Y = \frac{1}{2(1 - \frac{\ln r}{r-1})^2} + \frac{1}{2(\frac{r \ln r}{r-1} - 1)^2}, \tag{24}$$

and $r = \frac{\sigma_i^2}{\sigma_j^2}$, defined in Equation (19)*.*

From Theorem 2, the following observations can be made:

- Detection rate $P_d(Y)$ is an increasing function of sample size n. When $n \to \infty$, detection rate $P_d(Y) \to 100\%$. This implies that the more samples an adversary could collect, the higher detection rate by sample variance.
- As shown in the second part of Proof of Theorem 2 in Appendix B of our technical report [24], detection rate $P_d(Y)$ is an increasing function of r. That is, the smaller r, the lower detection rate. When $r = 1$, detection rate is 50%. This corresponds to the case when random delay with sufficiently large σ_p^2 is applied to the web traffic. This suggests that although the adversary may use a large size of sample to infer web sites by sample variance, extra random delay can make such an attack infeasible.

V. Countermeasures

In this section, we introduce countermeasures to mitigate the threat from the RTT based side channel. We start with a naive approach, followed by k-$means$ clustering based approach and $\mathbb{K}$-$Anonymity$ based approach.

A. Basic Countermeasure

According to Shannon's perfect secrecy principle, if the distribution of the RTT between the client and each web site is the same, the information gained from the attack approaches zero and the anonymity degree reaches maximum. Following this principle, we develop the statistical distribution based countermeasure. We select a web site, denoted as target web site from the pool of web sites to be protected. We manipulate the features of other web sites so that all web sites look like each other in terms of sample mean and sample variance.

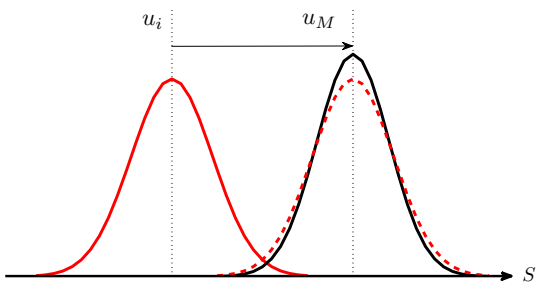

Fig. 4. Basic Approach

This basic countermeasure is discussed in details as follows. First, we choose a web site with the maximum sample mean from the pool of web sites. We presume the sample mean is normally distributed. Denote sample mean of the target web site as μ_M. Sample mean of the i^{th} web site is μ_i. Then we introduce delay to shift the RTT sample mean of the i^{th} web site close to the RTT sample mean of the target web site so that $\mu_M = \mu_i$, as illustrated in Figure 4. In this way, we increase the overlapping zone of these two distributions and increase the error rate in Equation (9). According to Theorem 1, the detection rate will decrease significantly since the two web sites have the same mean now.

However, the analysis of sample mean and sample variance in Section IV-C shows that even if the sample mean of all web sites is the same, sample variance of the RTTs may still disclose the web site. Theorem 2 also demonstrates that when r approaches 1, the web sites can not be identified. Recall from Equation (19), the variance ratio $r = \frac{\sigma_c^2+\sigma_p^2+\sigma_{w_i}^2}{\sigma_c^2+\sigma_p^2+\sigma_{w_j}^2}$. Consequently, we can increase sample variance σ_p^2 at the SSH server to achieve $r \to 1$. We can carefully select a maximum sample variance σ_M^2, which is larger than sample variance of all the web sites and manipulate the delay so that all web sites have the same variance. Therefore, by manipulating web traffic delay, all web sites will have the same distribution, $\mathcal{N}(\mu_M, \sigma_M^2)$.

We can use the squared Hellinger distance as a metric to estimate the distance between two normal distributions $\mathcal{P} \sim \mathcal{N}(\mu_i, \sigma_i^2)$ and $\mathcal{Q} \sim \mathcal{N}(\mu_M, \sigma_M^2)$,

$$\mathcal{D}_i = \quad ^2(\mathcal{P}, \mathcal{Q}) = 1 - \sqrt{\frac{2\sigma_M\sigma_i}{\sigma_M^2+\sigma_i^2}} e^{-\frac{1}{4}\frac{(\mu_M-\mu_i)^2}{\sigma_M^2+\sigma_i^2}}. \qquad (25)$$

The total distance is $\mathcal{D} = \sum_{i=1}^{N} \mathcal{D}_i$. In this way, the adversary cannot distinguish the web sites visited by users via RTT, and we can ensure the users can achieve the maximum anonymity degree, i.e., o N. However, this basic approach will introduce considerable delay into the web traffic, which significantly affects the web browsing performance. To resolve this problem, we introduce the following two enhanced approaches.

B. k-means Clustering based Approach

To balance the tradeoff between the performance and anonymity degree, we propose to use the k-$means$ algorithm to group web sites into k clusters. Web sites in one cluster have similar means. Denote t_1, t_2, ..., t_N as the feature set for N web sites, where t_i is a x-dimensional feature vector for the i^{th} web site. By using the k-$means$ clustering, we aim to partition the N web sites into k clusters $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k\}$ ($k \leqslant N$). We choose a feature vector for each cluster and will shape all web sites to have that feature vector. In our case, *<maximum sample mean, maximum sample variance>* is chosen as the feature vector of that cluster. The corresponding target distributions for these clusters are $\{\mathcal{N}(\mu_1', \sigma_1'), \mathcal{N}(\mu_2', \sigma_2'), \ldots, \mathcal{N}(\mu_N', \sigma_N')\}$. Therefore, for the i^{th} cluster, we manipulate all web sites to have a distribution of $\mathcal{N}(\mu_i', \sigma_i')\}$. Using Equation (25), we can derive the distance $\mathcal{D}'_i$ between each web site and the target web site for each cluster and obtain the total distance $\mathcal{D}' = \sum_{i=1}^{N} \mathcal{D}'_i$.

In the k-$means$ based countermeasure, traffic from a web site is shaped to have the distribution of its cluster, instead of a globe distribution as in the naive approach. Consequently, the network delay introduced by the k-$means$ clustering based countermeasure can be significantly reduced.

Algorithm 1 $\mathbb{K}$-*Anonymity* Algorithm

Require:
(a) $\mathcal{C}$, a set of clusters derived by using k-$means$ Clustering; (b) $\mathcal{C}_\mathcal{K}$, an empty set of clusters to satisfy k-$anonymity$ Clustering; (c) $\mathcal{C}_i$, certain cluster in $\mathcal{C}$; (c) $\mathcal{C}_m$, a cluster with the maximum number of web sites in $\mathcal{C}$; (d) $\mathcal{D}(\mathcal{C}_i, \mathcal{C}_j)$, the distance between $\mathcal{C}_i$ and $\mathcal{C}_i$; (e) $\mathcal{R}$, the remainder when site number is divided by $\mathbb{K}$.

Ensure: $\mathcal{C}$ Satisfies $\mathbb{K}$-*Anonymity*

while Not null in $\mathcal{C}$ **do**
  for $|\mathcal{C}_i| < \mathbb{K}$ **do**
    Search $\mathcal{C}$ to find $\mathcal{C}_j$ such that $\mathcal{D}(\mathcal{C}_i, \mathcal{C}_j)$ is minimum.
    Merger clusters $\mathcal{C}_i$ and $\mathcal{C}_j$
  end for
  for each cluster $\mathcal{C}_i$ such that $|\mathcal{C}_i| > 2\mathbb{K}$ **do**
    Partition the group into $\lceil \frac{|\mathcal{C}_i|}{\mathbb{K}} \rceil$ clusters to ensure that $\lfloor \frac{|\mathcal{C}_i|}{\mathbb{K}} \rfloor$ clusters have at least $\mathbb{K}$ web sites accord to minimum $\mathcal{C}_i$ and $\mathcal{C}_m$;
  end for
  while Exist certain cluster $|\mathcal{C}_i|$ is between $\mathbb{K}$ and $\mathbb{K}+\mathcal{R}$ **do**
    Transfer cluster $|\mathcal{C}_i|$ to $|\mathcal{C}_i| < \mathbb{K}$ and adjust $\mathcal{R}$;
  end while
  Partition $|\mathcal{C}_i|$ with maximum site number into 2 clusters to ensure that one has at least $\mathbb{K}$ web sites accord to minimum $\mathcal{C}_i$ and $\mathcal{C}_m$;
end while

C. $\mathbb{K}$-*Anonymity* based Approach

The astute reader would surely have noticed that one issue that we have seemingly ignored is the anonymity degree of

web sites in each cluster in the introduction of the k-$means$ strategy. If the number of web sites in a cluster is very small, the adversary can readily identify the set of web sites visited by a user. For example, if there is only one web site in a cluster, the attacker can definitively identify the web site. Therefore, we shall also ensure the anonymity degree for each cluster of web sites.

To address this issue, we introduce $\mathbb{K}$-$Anonymity$ to guarantee a minimum degree anonymity for all clusters. In particular, $\mathbb{K}$-$Anonymity$ is defined as if and only if there are at least $\mathbb{K}$ web sites in each cluster. According to the existing partition of the cluster via k-$means$, we use Algorithm 1 to achieve $\mathbb{K}$-$Anonymity$. We first define the distance between two clusters. Denote the target web site in $\mathcal{C}_i$ as $\mathcal{P}' \sim \mathcal{N}(\mu_i', (\sigma_i')^2)$, where μ_i' and $(\sigma_i')^2$ is the maximum sample mean and sample variance in $\mathcal{C}_i$, respectively. As such, let the target web site in $\mathcal{C}_j$ be $\mathcal{Q}' \sim \mathcal{N}(\mu_j', (\sigma_j')^2)$. Then we have the distance between $\mathcal{C}_i$ and $\mathcal{C}_j$ derived by

$$\mathcal{D}(\mathcal{C}_i, \mathcal{C}_j) = \quad ^2(\mathcal{P}', \mathcal{Q}'). \tag{26}$$

Then we use an efficient heuristic algorithm Algorithm 1 to achieve $\mathbb{K}$-$Anonymity$. The basic idea is that we choose the cluster $\mathcal{C}_i$, whose cardinality $|\mathcal{C}_i|$ is less than $\mathbb{K}$. Then we search a cluster $\mathcal{C}_j$, where $\mathcal{D}(\mathcal{C}_i, \mathcal{C}_j)$ is minimum. We then merge these two clusters. If the cardinality of the new cluster is larger than $2\mathbb{K}$, we partition the cluster into $\lfloor \frac{|\mathcal{C}_i|}{\mathbb{K}} \rfloor$ cluster. Eventually, we can get the final clusters that meet the $\mathbb{K}$-$Anonymity$ requirement.

VI. Evaluation

We have implemented the novel RTT based side-channel attack against encrypted web sites. In this section, we use real-world experiments to demonstrate the information leakage caused by the side-channel attack and feasibility and effectiveness of countermeasures proposed in Section V.

A. Experiment Setup

Figure 5 illustrates the experiment setup. We deployed a SSH sever in North America. Two other computers were deployed in Asia. One computer acts as a client and is connected to a wireless access point. The mimic attacking computer sniffs the timing of packets into and out of the client. All computers run Fedora Core 14. The version of OpenSSH [25] at both SSH server and client is $5.8p2$. The web browser is Firefox 3.6.17. We configure Firefox not to cache the data. The latest Adobe Flash plugin [26] is installed at the browser.

B. Detection Rate

To validate the effectiveness of the side channel attack, we target the top 100 web sites in the Alexa rankings [27]. Since the distribution of popularity of the web sites is a *Zipf* distribution, i.e., 80% users visit 20% web sites, this set of web sites are sufficient to demonstrate the potential threat from the network delay based side-channel. To access the web sites at the client side, we modify *Pagestats* [28] and drive Firefox to visit these web sites automatically. The libpcap 1.1.1 library and TCPdump 4.1.1 are used to capture packets. The client automatically accesses each web site one by one and dumps traffic from a web site through TCPdump. The client fetched each web site for 100 times and the experiments lasted for around one month. By using the approach in Section III-A, we can calculate RTTs between each of the top 100 web sites and the client, and establish the profiles for these web sites.

To validate the damage of the side-channel attack, we use data collected in odd days as the training set and data collected in even days as the test set. We use the *Weka* toolkit [29] to implement the classification in Section III-C, with Gaussian kernel density estimation. When a sample is collected, we will calculate sample mean and sample variance and generate a table of probabilities of how much the collected sample matches each web site in the trained profiles. We adopt the λ-$identifiability$ [4] for a classification instance. Given λ, the classification is deemed successful if the classification probability corresponding to the real web site falls into the top λ probabilities in the derived propagability table. Otherwise, the classification fails.

Figure 6 illustrates the detection rate in terms of different i5i5

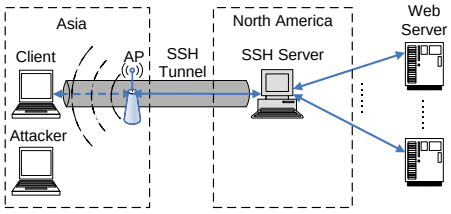

Fig. 5. Experiment Setup

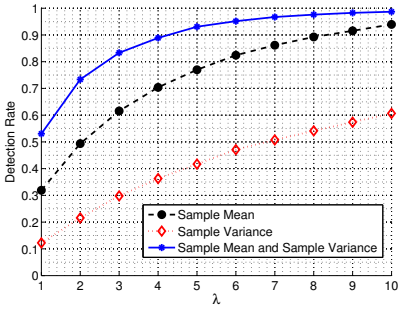

Fig. 6. Detection Rate v.s. λ for the Three Features: Sample Mean, Sample Variance, and the Tuple

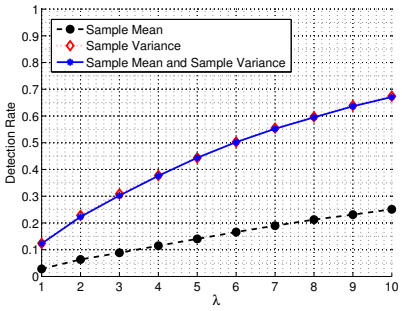

Fig. 7. Detection Rate v.s. λ for Basic Countermeasure Manipulating RTT Mean

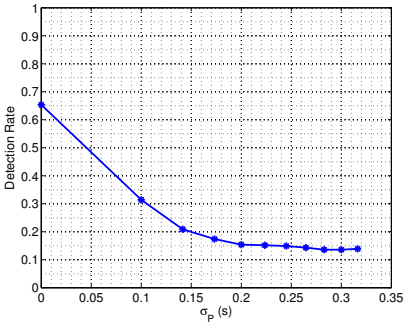

Fig. 8. Detection Rate for Manipulating RTT Variance σ_p^2 with $\lambda = 10$

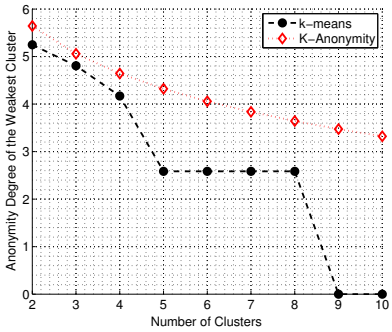

Fig. 9. Anonymity Degree v.s. Number of Clusters

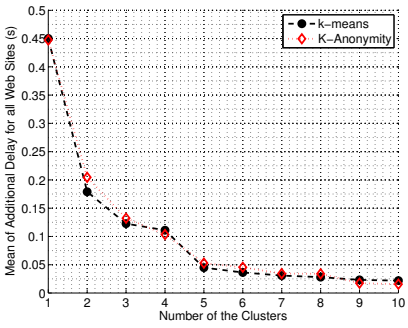

Fig. 10. Average Extra Delay v.s. Number of Clusters

the adversary can readily identify this web site from the rest. As we can see, the k-*means* clustering fails to preserve user privacy in this case. In comparison, the $\mathbb{K}$-*Anonymity* based strategy can still prevent the adversary from differentiating web sites from each cluster. When the number of clusters is 10, the $\mathbb{K}$-*Anonymity* based strategy achieved an anonymity degree of more than 3 bits (corresponds to 8 websites).

Figure 10 shows the average of *extra delay* introduced into all web sites. It can be observed when the clustering approach (number of the clusters > 1) is used, the delay is significantly reduced. In addition, we can see that delay introduced by using the $\mathbb{K}$-*Anonymity* based approach is similar to the k-*means* based countermeasure while the former achieves a guaranteed anonymity degree. These observations verify that the $\mathbb{K}$-*Anonymity* based countermeasure can tackle the tradeoff between the performance and anonymity degree for clusters of web sites.

VII. Conclusion

In this paper, we investigated a novel side-channel attack via RTT between a victim user and encrypted web sites. We define detection rate as the probability that an adversary correctly recognizes web sites accessed by a victim user. We conducted careful analysis of different statistics of RTTs and found that sample mean and sample variance of web traffic RTTs can disclose web sites being accessed. We theoretically analyzed the detection rate and derived closed-form formulae. To defend against this attack, we proposed three countermeasures: the basic approach shaping all web traffic to have a similar RTT in terms of mean and variance, the k-*means* strategy grouping web sites with similar RTT mean and variance into the same cluster in order to address the performance issue of the basic approach, and the k-anonymity strategy to guarantee a minimum degree of anonymity for web sites in one cluster. We conducted extensive experiments to evaluate the severeness of the side channel-attack and feasibility and effectiveness of the countermeasures. Our experiments demonstrate that the $\mathbb{K}$-*Anonymity* based countermeasure can effectively balance the tradeoff between the performance and information leakage. Our empirical results match our theoretical analysis very well.

Acknowledgement

This work was supported in part by National Key Basic Research Program of China under Grants 2010CB328104, NSFC (China) under Grants 60903162, 60903161, 61070158, 61070161, 61003257, China Specialized Research Fund for the Doctoral Program of Higher Education under Grants 200802860031, Jiangsu Provincial Natural Science Foundation of China under Grants BK2008030, Jiangsu Provincial Key Laboratory of Network and Information Security under Grants BM2003201, and Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under Grants 93K-9, and USA NSF grants 0942113, 0958477, 0943479 and CNS-1117175 and the Army Research Laboratory under grant W911NF-11-1-0193. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of those sponsors.

References

[1] H. Roberts, E. Zuckerman, J. York, R. Faris, and J. Palfrey, "2010 circumvention tool usage report," http://cyber.law.harvard.edu/sites/cyber.

law.harvard.edu/files/2010_Circumvention_Tool_Usage_Report.pdf, 2010.

[2] Q. Sun, D. R. Simon, Y.-M. Wang, W. Russel, V. N. Padmanabhan, and L. Qiu, "Statistical identification of encrypted web browsing traffic," in *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, May 2002.

[3] G. D. Bissias, M. Liberatore, D. Jensen, and B. N. Levine, "Privacy vulnerabilities in encrypted http streams," in *Proceedings of the 5th Privacy Enhancing Technologies Workshop (PET)*, May 2005.

[4] M. Liberatore and B. N. Levine, "Inferring the source of encrypted http connections," in *Proceedings of the ACM conference on Computer and Communication Security (CCS)*, October 2006.

[5] D. Herrmann, R. Wendolsky, and H. Federrath, "Website fingerprinting: attacking popular privacy enhancing technologies with the multinomial naïve-bayes classifier," in *Proceedings of the 2009 ACM workshop on Cloud computing security (CCSW)*, September 2009.

[6] L. Lu, E.-C. Chang, and M. C. Chan, "Website fingerprinting and identification using ordered feature sequences," in *Proceedings of the 15th European conference on Research in Computer Security (ESORICS)*, September 2010.

[7] S. Chen, R. Wang, X. Wang, and K. Zhang, "Side-channel leaks in web applications: a reality today, a challenge tomorrow," in *Proceedings of the 31st IEEE Symposium on Security and Privacy (S&P)*, May 2010.

[8] C. V. Wright, S. E. Coull, and F. Monrose, "Traffic morphing: An efficient defense against statistical traffic analysis," in *Proceedings of the Network and Distributed Security Symposium (NDSS)*, February 2009.

[9] X. Luo, P. Zhou, E. W. W. Chan, W. Lee, R. K. C. Chang, and R. Perdisci, "Httpos: Sealing information leaks with browser-side obfuscation of encrypted flows," in *Proceedings of the 18th Annual Network and Distributed System Security Symposium (NDSS)*, February 2011.

[10] Y. Zhu, X. Fu, B. Graham, R. Bettati, and W. Zhao, "On flow correlation attacks and countermeasures in mix networks," in *Proceedings of Workshop on Privacy Enhancing Technologies (PET)*, May 2004.

[11] B. N. Levine, M. K. Reiter, C. Wang, and M. Wright, "Timing attacks in low-latency mix-based systems," in *Proceedings of Financial Cryptography (FC)*, February 2004.

[12] D. X. Song, D. Wagner, and X. Tian, "Timing analysis of keystrokes and timing attacks on ssh," in *Proceedings of 10th USENIX Security Symposium*, August 2001.

[13] Q. X. Sun, D. R. Simon, Y. Wang, W. Russell, V. N. Padmanabhan, and L. L. Qiu, "Statistical identification of encrypted web browsing traffic," in *Proceedings of IEEE Symposium on Security and Privacy (S&P)*, May 2002.

[14] C. V. Wright, L. Ballard, F. Monrose, and G. M. Masson, "Language identification of encrypted voip traffic: Alejandra y roberto or alice and bob?" in *Proceedings of the 16th Annual USENIX Security Symposium (Security)*, August 2007.

[15] C. V. Wright, L. Ballard, S. E. Coull, F. Monrose, and G. M. Masson,