Time and Place: Robustness of a Traffic Analysis Attack Against Web Traffic

Saman Feghhi, Douglas J. Leith
School of Computer Science and Statistics, Trinity College Dublin
Email: {feghhis,doug.leith}@tcd.ie

***Abstract*—We consider a timing-only based attack against encrypted and padded web traffic. The attacker can collect training data, but only over a different connection from that against which the attack is directed. This is a significantly easier to perform attack than one which depends on training data collected over the victim link. We demonstrate than an attacker can infer the correct web page >87% of the time when the training data is collected at a distance of up to 25 km from the victim, provided that the type of link is similar *e.g.* if the victim link uses a cable modem then the training data should be measured over a cable modem link. We also investigate the impact of distance in time between when the training data is collected and when the attack is performed.**

I. Introduction

In this paper we consider an attacker of the type illustrated in Figure 1. A client A browses the web over an encrypted connection, *e.g.* a VPN, and the attacker can observe the encrypted traffic. It is assumed that the encrypted packets are padded to be the same size so that the packet size reveals no information about the nature of the traffic. In this case the only information available is the timing pattern of the traffic traversing the link. The attacker's objective is to guess, with high probability of success, the web sites being visited by the victim. As the attacker is only relying on timestamp information, this attack is impervious to the existing defences which obscure packet size information.

To assist with the attack the attacker can, of course, themselves fetch web pages of interest and record the packet timings for use as training data (against which the victims data can be compared). An attack of this type making use of training data collected over the victims link was previously considered in [3] and it was demonstrated that the web pages being browsed could indeed be accurately inferred with high probability, achieving mean success rates in excess of 90% for both wired and wireless traffic. However, it is often relatively difficult to collect such training data in an unobstrusive manner since it takes time and creates a significant volume of traffic over the victims link. In this paper we therefore consider the feasibility of an attack based on training data collected over a link which is different from the victims. This might be located at a neighbouring house or office, but of more interest is whether training data can usefully be collected at a location much further away *e.g.* at a location within the same city as the victim but otherwise not in their vicinity. Such an attack would clearly be relatively easy to carry out and of significant concern.

This work was supported by Science Foundation Ireland under Grant No. 11/PI/1177.

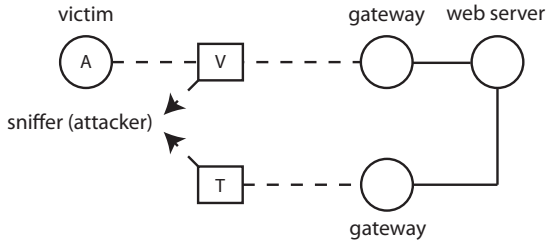


Fig. 1: Schematic illustrating attacker of the type considered. The targeted victim is connected to the Internet via an encrypted tunnel (ssh, SSL, IPSec *etc.*) shown by dashed lines. The attacker can measure the timing of the victims packets traversing the tunnel at point V in the uplink direction. In addition, the attacker can make their own measurements over a second link T, different from that used by the victim.

We demonstrate that an attack which infers the correct web page >87% of the time is feasible even when the training data is collected at a distance of up to 25 km from the victim, provided that the type of link is similar *e.g.* if the victim link uses a cable modem then the training data should be measured over a cable modem link.

We further investigate the impact of distance in time between when the training data is collected and when the attack is performed. As might be expected, as the time elapsed increases the accuracy of the attack falls. However, the rate of decrease is surprisingly small: up to two months between collection of the training data and performance of the attack around 80% of web pages are estimated with >70% accuracy. Eventually, after seven months the success rate falls such that only 30% of web pages are estimated with >70% accuracy.

II. Related Work

The general topic of traffic analysis attacks against encrypted web traffic has been the subject of much interest, and a fairly large body of literature exists. However, almost all of this is concerned with attacks based primarily on packet size and count. To the best of our knowledge all fingerprinting attacks involve use of training data collected at the same link on which the attack is to be performed. However [4] considers a different type of attack where a remote adversary living in a different network than the victim, attempts to infer contents of their

traffic by analysing the round-trip times (RTTs) of ICMP echo messages sent to the victim. An exception to attacks based primarily on packet size/count is the attack bases solely on timing which was previously reported by the authors in [3], and upon which the present paper builds (in [3] the training and test data as collected at the same link and at almost the same time). Regarding web traffic attacks where the training data is collected at a different time from when the attack is performed, [1] considers an attack where the training samples are taken hourly over a period of three months. For 100 websites the accuracy of the attack using 24-hour training data is 23%, although this improves when the number of guesses are increased or the dataset is narrowed down to a small subset of the web sites. In [2] the training samples are collected using different platforms, but it is unclear whether the platforms are within the same network or if samples of the same web page are fetched over different platforms.

III. Preliminaries

A. Anatomy of a Web Page Fetch

When traffic is carried over an encrypted tunnel the packet source/destination addresses/ports and the payload are hidden. We also assume here that the tunnel pads the packets to be of equal size, so that packet size information is also concealed. An attacker sniffing traffic on the encrypted tunnel is therefore able only to observe the direction and timing of packets through the tunnel, *i.e.* to observe a sequence of pairs $\{(t_k, d_k)\}$, $k = 1, 2, \cdots$ where t_k is the time at which the k'th packet is observed and $d_k \in \{-1, 1\}$ indicates whether the packet is travelling in the uplink/downlink direction. Since it will be sufficient to mount an effective attack, we will assume a weaker attacker that can only observe the timestamps $\{t_k\}$, $k \in K_{up} := \{\kappa \in \{1, 2, \cdots\} : d_\kappa = -1\}$ of the uplink traffic.

To gain insight into the nature of packet timestamp sequences, it is helpful to consider the process of fetching a web page in more detail. When fetching a web page the client browser opens a TCP connection with the server indicated by the URL and issues an HTTP GET or POST request to which the server then replies. As the client parses the server response it issues additional GET/POST requests to fetch embedded objects (images, css, scripts *etc.*). These additional requests may be to different servers from the original request (*e.g.* when the object to be fetched is an advert or is hosted in a separate content-delivery network), in which case the client opens a TCP connection to each new server in order to issue the requests. Fetching of these objects may in turn trigger the fetching of further objects. We make the following observations:

1) *Connection to third-party servers*. Fetching an object located on a third-party server requires the opening of a new TCP connection to that server, over which the HTTP request is then sent. The TCP connection handshake introduces a delay (of at least one RTT) and since the pattern of these delays is related to the web page content it can potentially assist in identifying the web page.
2) *Pipelining of requests*. Multiple objects located on the same server lead to several GET/POST requests being sent to that server, one after another. Due to the dynamics of TCP congestion control, this burst of back-to-back requests can affect the timing of the response packets in a predictable manner that once again can potentially assist in identifying the web page.
3) *Asynchronous requests*. Dynamic content, *e.g.* pre-fetching via AJAX, can lead to update requests to a server with large inter-arrival times that can potentially act as a web page signature.
4) *Connection closing*. When a web page fetch is completed, the associated TCP connections are closed. A FIN/FINACK/ACK exchange closes each connection and this burst of packets can have quite distinctive timing which allows it to be identified. Since the number of connections is related to the number of distinct locations where objects in the web page are stored, it changes between web pages.

Following the approach in [3], we use timing features such as these, which vary depending upon the web page fetched, to create a timing signature which allows us to identify which web page is being fetched based on timing data only.

B. Classifying Measured Timestamp Sequences

Suppose we have two sequences of packet timestamps $\boldsymbol{t} := \{t_i\}$, $i = 1, 2, \cdots, n$ and $\boldsymbol{t'} := \{t'_j\}$, $j = 1, 2, \cdots, m$. Note that the sequence lengths n and m are *not* assumed to be the same. To proceed we need to define an appropriate measure of the distance between such sequences.

1) Derivative Dynamic Time Warping: To classify time sequences of each web page we need a measure of the distance between packet sequences which is insensitive to the types of distortion introduced by the network, so that the distance between packet streams t and t' associated with fetches of the same web page at different times is measured as being small, and ideally the distance between fetches of different web pages is measured to be large. We use a variant of Dynamic Time Warping (DTW) [5]. DTW aims to be insensitive to differences between sequences which are due to stretching/compressing of time and so can be expected to at least partly accommodate the effects of changes in download rate, queueing delay *etc.* and distortions due to using a shared wireless channel.

We define a warping path $\boldsymbol{p}$ to be a sequence of pairs, $\{(p_k^i, p_k^j)\}$, $k = 1, 2, \cdots, l$ with $(p_k^i, p_k^j) \in V := \{1, \cdots, n\} \times \{1, \cdots, m\}$ satisfying boundary conditions $p_1^i = 1 = p_1^j$, $p_l^i = n$, $p_l^j = m$ and step-wise constraints $(p_{k+1}^i, p_{k+1}^j) \in V_{p_k^i, p_k^j} := \{(u, v) : u \in \{p_k^i, p_k^i + 1\} \cap \{1, \ldots, n\}, v \in \{p_k^j, p_k^j + 1\} \cap \{1, \ldots, m\}\}$, $k = 1, \cdots, l-1$. That is, a warping path maps points from one timestamp sequence to another such that the start and end points of the sequences match (due to the boundary conditions) and the points are monotonically increasing (due to the step-wise constraints). This is illustrated schematically in Figure 2, where the two timestamp sequences to be compared are indicated to the left and above the matrix and the bold line indicates an example warping path.

Let $P_{mn}^l \subset V^l$ denote the set of all warping paths of length l associated with two timestamp sequences of length n and m respectively, and let $C_{\boldsymbol{t},\boldsymbol{t'}}(\cdot) : P_{mn}^l \to \mathbb{R}$ be a

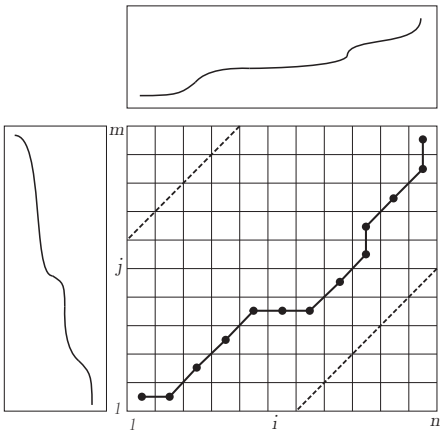


Fig. 2: Illustrating a warping path. The dashed lines indicate the warping window.

cost function so that $C_{\boldsymbol{t},\boldsymbol{t}'}(\boldsymbol{p})$ is the cost of warping path $\boldsymbol{p} \in P^l_{mn}$. Our interest is in the minimum cost warping path, $\boldsymbol{p}^*(\boldsymbol{t},\boldsymbol{t}') \in \arg\min_{\boldsymbol{p}\in P^l_{mn}} C_{\boldsymbol{t},\boldsymbol{t}'}(\boldsymbol{p})$. In DTW the cost function has the separable form $C_{\boldsymbol{t},\boldsymbol{t}'}(\boldsymbol{p}) = \sum_{k=1}^{l} c_{\boldsymbol{t},\boldsymbol{t}'}(p^i_k, p^j_k)$ where $c_{\boldsymbol{t},\boldsymbol{t}'} : V \to \mathbb{R}$, in which case optimal path $\boldsymbol{p}^*(\boldsymbol{t},\boldsymbol{t}')$ be efficiently found using the backward recursion,

$$(p^i_k, p^j_k) \in \arg \min_{(p^i,p^j)\in V_k} C_{k+1} + c_{\boldsymbol{t},\boldsymbol{t}'}(p^i, p^j) \tag{1}$$

$$C_k = C_{k+1} + c_{\boldsymbol{t},\boldsymbol{t}'}(p^i_k, p^j_k) \tag{2}$$

where $V_k = (p^i, p^j) \in \{(u,v) : (p^i_{k+1}, p^j_{k+1}) \in V_{u,v}\}$, $k = l-1, l-2, \cdots$ and initial condition $C_l = c_{\boldsymbol{t},\boldsymbol{t}'}(n,m)$. When there is more than one optimal solution at step (1), we select (p^i_k, p^j_k) uniformly at random from amongst them.

One common choice of element-wise cost is the Euclidean norm $c_{\boldsymbol{t},\boldsymbol{t}'}(p^i, p^j) = (t_{p^i} - t'_{p^j})^2$. However, to improve robustness to noise on the timestamp values (in addition to misalignment of their indices), following [5] we instead use the following element-wise cost

$$c_{\boldsymbol{t},\boldsymbol{t}'}(p^i, p^j) = (D_{\boldsymbol{t}}(p^i) - D_{\boldsymbol{t}'}(p^j))^2 \tag{3}$$

where $D_{\boldsymbol{t}}(i) = \frac{(t_i - t_{i^-}) + (t_{i^+} - t_{i^-})}{2}$, $i^- = \max\{i-1, 1\}$ and $i^+ = \min\{i+1, |\boldsymbol{t}|\}$. Observe that $D_{\boldsymbol{t}}(i)$ is akin to the derivative of sequence $\boldsymbol{t}$ at index i. Further, we constrain the warping path to remain within windowing distance w of the diagonal (*i.e.* within the dashed lines indicated on Figure 2) by setting $C(\boldsymbol{p}) = +\infty$ for paths $\boldsymbol{p} \in P^l_{mn}$ for which $|p^i_k - p^j_k| > \max\{w \min\{n,m\}, |m-n|\}$ for any $k \in \{1, \cdots, l\}$. For the rest of this paper we use $w = 0.2$.

2) F-Distance Measure: Given two timestamp sequences, the warping path is a mapping between them. With reference to Figure 2, sections of the warping path which lie parallel to the diagonal correspond to intervals over which the two sequences are well matched. Sections of the warping path that are parallel to the x- or y-axes correspond to intervals over which the two sequences are poorly matched. This suggests using the fraction of the overall warping path which is parallel to the x- or y-axes as a distance measure, referred to as the F-distance.

In more detail, let $\boldsymbol{p} = \{(p^i_k, p^j_k)\}$, $k = 1, \cdots, l$ be a derivative DTW warping path relating timestamp sequences $\boldsymbol{t}$ and $\boldsymbol{t}'$, obtained as described in the previous section. We partition the warping path into a sequence of subpaths within each of which either p^i_k or p^j_k remain constant and we count the subpaths which are longer than one. Formally, define $\kappa_1 := 0 < \kappa_2 < \cdots < \kappa_{r-1} < \kappa_r := l$ such that for each $s = 1, \cdots, r-1$ (i) either $p^i_{k_1} = p^i_{k_2}$ $\forall k_1,\ k_2 \in \{\kappa_s+1, \cdots, \kappa_{s+1}\}$ or $p^j_{k_1} = p^j_{k_2}$ $\forall k_1, k_2 \in \{\kappa_s+1, \cdots, \kappa_{s+1}\}$ and (ii) either $\kappa_{s+1} = l$ or condition (i) is violated for some $k_1, k_2 \in \{\kappa_s, \cdots, \kappa_{s+1}+1\}$ *i.e.* each subsequence is maximal. Note that $p^i_k \neq p^j_k$ for all $k = 1, \cdots, l$ (due to warping path step-wise constraints) and so in condition (i) it is not possible for both p^i_k and p^j_k to be constant. The F-distance measure between timestamp sequences $\boldsymbol{t}$ and $\boldsymbol{t}'$ is defined as:

$$\phi(\boldsymbol{t},\boldsymbol{t}') := \frac{\sum_{s\in\{1,\ldots,r-1\}} \kappa_{s+1} - \kappa_s}{\kappa_{s+1}-\kappa_s>1} \Big/ (n+m) \tag{4}$$

where κ_s, $s = 1, \cdots, r$ are the constant subsequences in minimal warping path $\boldsymbol{p}^*(\boldsymbol{t},\boldsymbol{t}')$. It can be seen that $\phi(\boldsymbol{p})$ takes values in interval $[0,1]$, and is 0 when sequences $\boldsymbol{t}$ and $\boldsymbol{t}'$ are identical (in which case the warping path $\boldsymbol{p}$ lies on the diagonal in Figure 2).

3) k-Nearest Neighbours Classification: A k-Nearest Neighbours approach was used for classification. For each web page i we sort the measured timestamp sequences $\boldsymbol{t}' \in T_i$ used for training in ascending order of sum-distance $\sum_{\boldsymbol{t}\in T_i} \phi(\boldsymbol{t},\boldsymbol{t}')$ and select the top 3 to use as exemplars to represent the web page. When presented with a new timestamp sequence, its distance to the exemplars for all of the training web pages is calculated and these distances are sorted in ascending order. Classification is then carried out by majority vote amongst the top k. All the measurements in this paper are for $k = 5$.

IV. De-anonymizing Web Fetches Using Training Data From Different Locations

A. Measurements

We collected packet trace timing measurements at a variety of locations in Dublin and Maynooth (a town about 20 km west of Dublin) in Ireland.

1) Hardware/Software Setup: The target machine is a Sony VGN-Z11MN laptop with an Intel core 2 duo 2.26GHz CPU and 4GB of memory. It is running Ubuntu Linux 14.04 LTS Precise. A `watir-webdriver` script on Firefox 36.0 was used to perform the web page fetches and *tcpdump* to record the timestamps and direction (uplink/downlink) of all packets traversing a tunnel although only packet timestamps on the uplink are actually used.

2) Web Pages: At each measurement time and location we fetched the home pages of each of the top Irish health, financial and legal web sites as ranked by www.alexa.com under its Regional/Europe/Ireland category. We prune the pages that fail to load and then for each of the top 100 sites we carry out 100

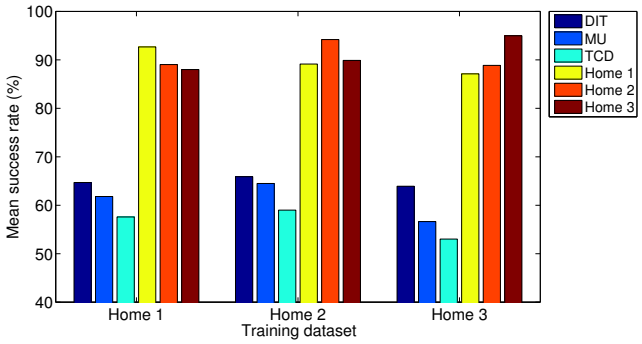


Fig. 3: Performance of attack against different locations (specified in legend) when using training data from a home connection (specified on x-axis).

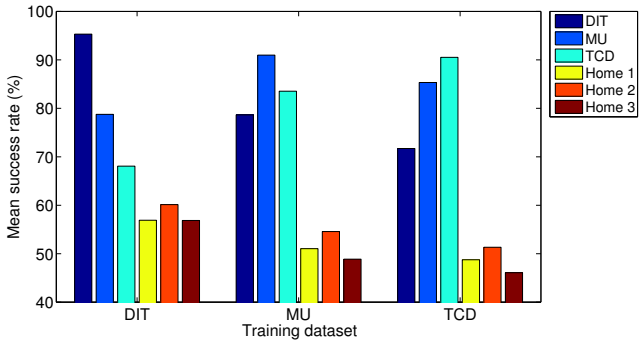


Fig. 4: Performance of attack when using training data from a university campus connection.

fetches of the index page yielding a total of 10,000 individual web page fetches in the data set collected at each time and location. Collection of these 10,000 pages was carried out over a period 4-5 days (so we collected more than 24 days of data across 6 locations). In these datasets the browser cache is flushed between each fetch so that the browser always starts in a fresh state. Fetches for different web sites are interleaved so that the measurements collected for a given web site are evenly spread over the duration of the measurement campaign at a particular location, to try to avoid bias due to changes in network conditions with time of day.

3) Place: We collected measurement datasets at 6 locations: (i) 3 university campus' namely, Trinity College Dublin, Dublin Institute of Technology and Maynooth University during 7-11 Jul, 24-28 Jun and 30 Jun-4 Jul 2015 respectively; (ii) 3 households, namely 2 in Dublin during 20-24 Jul and 5-9 Aug 2015, and one in Maynooth during 13-17 Jul 2015. For the campus measurements the target machine was connected via a gigabit ethernet LAN to the university network and then in turn to the Internet via a 10Gbps connection. For the household measurements the target machine is connected via ethernet to a cable modem and then to the Internet via a commercial Irish broadband network.

B. Results

Using the data measured at each location in turn as the training data, we evaluated the performance of the timing-only attack/classification method described in Section III-B when applied against the data measured at the remaining locations. The results obtained are summarised in Table I. This data is also visualised in Figures 3 and 4.

It can be seen that when training data is collected from the same location at which the attack is performed in (the diagonal entries in Table I) then the percentage of web pages which are correctly identified exceeds 90% at all locations. More interesting is that when measurements taken at a different home location are used as training data for an attack at another home location then the impact on the success rate is less than 5% *i.e.* more than 87% of web pages are successfully identified regardless of where the training data is collected, so long as it is collected at a home location. Note that Homes 2 and 3 are spaced well apart (some 15km) in Dublin city while Home 1 is in a separate town some 20 km outside Dublin. While it might seem plausible that the differences in timing and network conditions would significantly degrade the performance of the attack, our data indicates otherwise and establishes that a successful timing-based attack based on remotely collected training data is indeed possible.

Observe, however, that it is important that training data is collected over a similar network connection to that being attacked. It can be seen from Figure 3 that when data collected on a home connection is used for training then the success rate for the attack against university campus connection falls to between 53% and 65%. Note that this is still quite a high success rate – the success rate expected when selecting a web page uniformly at random from the set of 100 pages studies is only 1% – but is significantly lower than the $>$87% observed when using a home connection for training.

Similar behaviour is observed for attacks against a home connection. It can be seen from Table I that when another campus connection is used for training then the success rates for university connections remain above about 70% whereas for the attack against a home connection the success rate falls to between 46% and 60%. This is also evident from Figure 4.

Observe that the success rate when using a remote university connection for training in an attack against a university connection is lower ($68 - 85\%$) than when using a remote home connection for training in an attack against a home connection ($87 - 89\%$). We inspected the data in more detail to try to better understand this phenomenon. Based on this analysis we found that a fairly strong correlation exists between the classification accuracy and the difference in the maximum throughput measured at the training and test locations, see Figure 5. Although this requires further investigation, it seems likely that consistent differences in network traffic load at the various university campus' is the source of the behaviour observed – we note that the campus' are significantly different

Training \ Test	DIT	MU	TCD	Home 1 (Maynooth)	Home 2 (Dublin)	Home 3 (Dublin)
DIT	95.31%	78.76%	68.08%	56.91%	60.13%	56.87%
MU	78.65%	90.98%	83.53%	51.04%	54.57%	48.87%
TCD	71.71%	85.34%	90.53%	48.75%	51.34%	46.09%
Home 1 (Maynooth)	64.69%	61.81%	57.9%	92.68%	89.04%	88%
Home 2 (Dublin)	65.89%	64.49%	58.99%	89.14%	94.20%	89.89%
Home 3 (Dublin)	63.93%	56.64%	53.03%	87.11%	88.86%	94.98%

TABLE I: Summary of the measured success rate of the timing-only attack vs the location used to collect training data and the location used to collect test data (36 pairs of tests at 6 locations).

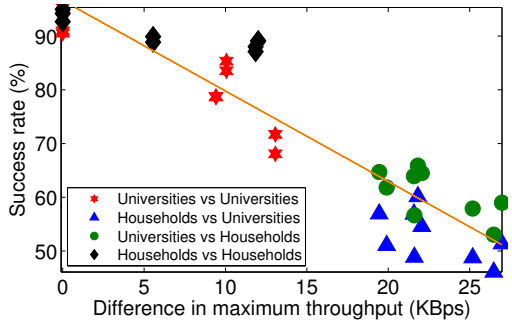


Fig. 5: Classification performance vs the difference (root mean square distance) in maximum throughput measured at training and test locations. Data is shown for the maximum uplink throughput but a similar correlation is also observed for the downlink and aggregate uplink/downlink throughput.

in size and number of users[1]. For home connections, the scale of differences in traffic load might be expected to be lower both because the raw network bandwidth is less and because there are fewer users.

V. Impact of Elapsed Time Lapse on Performance

The foregoing results establish the feasibility of collecting training data at a different location from that where the attack is performed. In this section we investigate the manner in which the elapsed time between collection of training data and performance of the attack affects its success. We expect the success of an attack to degrade as the time difference increases since many web pages are dynamic in nature and even small changes in news feeds or replacing images with new ones may well eventually accumulate and change the web page so much that the training data becomes less useful.

A. Measurements

1) University Campus: Using the same hardware/software setup and set of web pages as detailed previously, we collected measurements at Maynooth University over the periods Oct-Dec 2014, 16-20 Apr 2015 and from 30 Jun to 4 Jul 2015 and from Trinity College Dublin over the periods of 7-11 Jul, 1-5 Oct and 6-10 Oct 2015. The samples of each website are fetched once every hour over a period of 5 days except for the first Maynooth dataset where samples are fetched consecutively (this was part of an early measurement campaign).

2) Femtocell: We also collected data over a third type of network connection, namely a wireless femtocell. A femtocell is an eNodeB cellular base station with a small physical footprint (similar to a Wi-Fi access point) and limited cell size (typically about 30m radius) that is intended to improve cellular coverage indoors, filling in coverage holes and improving download rates, while also offloading traffic from the macrocell network. The femtocell used is a commercial Alcatel-Lucent 9361 Home Cell V2-V device, with traffic to/from the femtocell is backhauled over the campus ethernet connection at Maynooth University. A monitor computer running on an AMD Athlone 64 X2 Dual Core Proc 5000+ CPU and 4GB memory sniffs this encrypted traffic via a NetGear EN 108 TP Ethernet hub and logs all packets. The client machine is as for the other tests, but now equipped with a Huawei K3770 HSPA USB Broadband Dongle to connect wirelessly to the internet via the Femotcell. Datasets were collected during 11-16 Feb and 6-11 May 2015.

B. Results

Using the data measured in Maynooth University and Trinity College Dublin at each time period in turn as the training data, we evaluated the performance of the timing-only attack/classification when applied against the data measured at the remaining time periods. The results obtained are summarised in Table II. It can be seen that when the training data is collected around the same time as the attack is performed (the diagonal entries in the table), the success rate of the attack is $>$90%, in line with the results presented in the previous section. However, the success rate falls as the elapsed time between collection of the training data and performance of the attack is increased, as might be expected. The success rate lowers to 71% after 2 months and to 40% and 32% after 6 and 8 months respectively. Note that while 71% is significantly lower than the 90% success rate obtained when using contemporaneous training data, it is still a relatively high value that would likely be of concern. That is, our data indicates that a viable attack may be carried out even with a fairly large elapsed time of 2 months between the collection of training data and performance of the attack.

Fig 6 shows this data in more detail, plotting the distribution of the success rate over the set of 100 web pages studied vs the elapsed time between training and attack. It can be seen even when the average success rate over all web

[1]Recall that we collected data for each web page evenly over each measurement campaign to control for differences in traffic load over the course of a day, and we find no evidence correlation between success rate and time of day.

Training \ Test	MU 1	MU 2	MU 3
MU 1	94.98%	40.31%	32.5%
MU 2	40.77%	90.74%	71.71%
MU 3	34.37%	71.91%	90.98%

(a) Maynooth (Ethernet)

Training \ Test	FM 1	FM 2
FM 1	91.83%	43.79%
FM 2	47.03%	89.65%

(b) Maynooth(Femtocell)

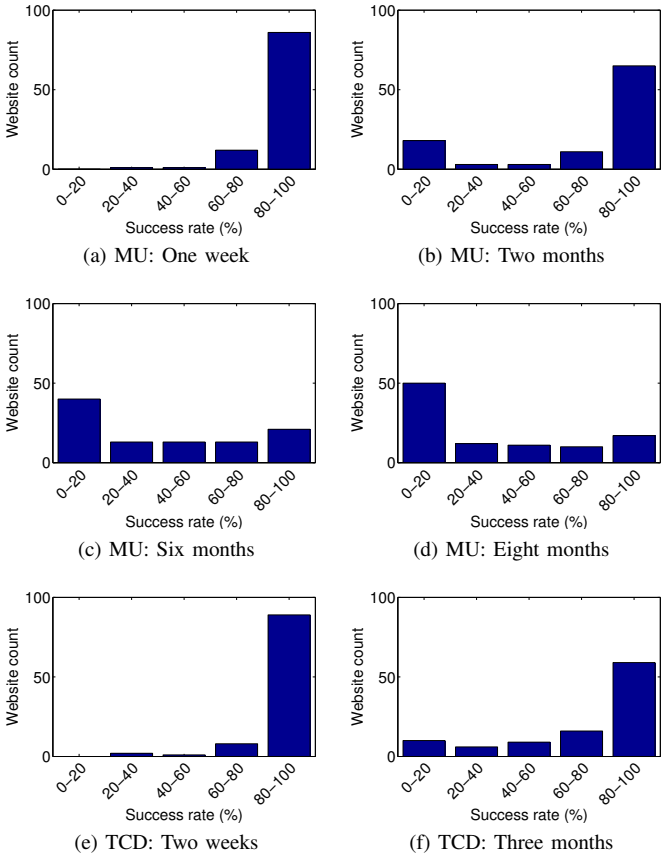
Training \ Test	TCD 1	TCD 2	TCD 3
TCD 1	90.53%	76.31%	76.03%
TCD 2	72.62%	92.42%	90.56%
TCD 3	72.68%	91.38%	93.42%

(c) TCD (Ethernet)

TABLE II: Measured success rate of the timing-only attack for Maynooth University ethernet and femtocell and Trinity College Dublin ethernet from measurements taken over different time periods (MU 1 = Oct-Dec 2014, MU 2 = 16-20 Apr 2015, MU 3 = 30 Jun to 4 Jul 2015, TCD 1 = 7-11 Jul 2015, TCD 2 = 1-5 Oct 2015, TCD 3 = 6-10 Oct 2015, FM 1 = 11-16 Feb 2015, FM 2 = 6-11 May 2015).

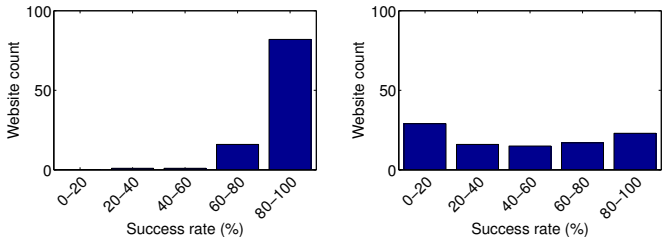
pages falls vs elapsed time (see Table II), a significant fraction of the individual web pages continue to be identified with high accuracy. We investigated this behaviour in more detail manually and find that for the web pages which experience a high drop in success rate either (i) the web site is down (returns 404 not found) in one test but not another, or (ii) the web page has changed significantly (the number of GET requests and objects is significantly different).

For comparison, the results obtained using the measurements taken over a femtocell link are shown in Table IIb. It can be seen that after three months have elapsed the success rate of the attack falls to less than 50% *i.e.* the degradation occurs somewhat more quickly than with the campus ethernet measurements. Figure 7 shows the corresponding distribution of success rates over the set of 100 web pages.



(a) MU: One week (b) MU: Two months (c) MU: Six months (d) MU: Eight months (e) TCD: Two weeks (f) TCD: Three months

Fig. 6: Distribution of success rates for individual web pages vs the elapsed time between collection of the training data and performance of the attack. Datasets are collected from Maynooth University and Trinity College Dublin.



(a) One week (b) Three months

Fig. 7: Distribution of success rates of individual web pages for different time gaps between training and test datasets. Datasets are collected from a femtocell channel running over Maynooth University campus internet.

VI. Conclusions

In this paper we consider a timing-only based attack against encrypted and padded web traffic. The attacker can collect training data, but only over a different connection from that against which the attack is directed. This is a significantly easier to perform attack than one which depends on training data collected over the victim link. We demonstrate than an attacker can infer the correct web page >87% of the time when the training data is collected at a distance of up to 25 km from the victim, provided that the type of link is similar *e.g.* if the victim link uses a cable modem then the training data should be measured over a cable modem link. We also investigate the impact of distance in time between when the training data is collected and when the attack is performed and show that a success rate of >70% can be achieved with a gap of as much as 2 months between training and attack.

Acknowledgement

The assistance of the Hamilton Institute, Maynooth University in facilitating data collection is gratefully acknowledged. Also that of our TCD colleague Mohammad Karzand.

REFERENCES

[1] G. D. Bissias, M. Liberatore, D. Jensen, and B. N. Levine. Privacy vulnerabilities in encrypted HTTP streams. *Lecture notes in computer science*, 3856:1, 2006.

[2] X. Cai, X. Ch. Zhang, B. Joshi, and R. Johnson. Touching from a Distance: Website Fingerprinting Attacks and Defenses. In *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, CCS '12, pages 605–616, New York, NY, USA, 2012. ACM.

[3] S. Feghhi and D. J. Leith. A Web Traffic Analysis Attack Using Only Timing Information. *arXiv*, abs/1410.2087, 2014.

[4] X. Gong, N. Borisov, N. Kiyavash, and N. Schear. Website detection using remote traffic analysis. In *Privacy Enhancing Technologies*, pages 58–78. Springer, 2012.

[5] E. J. Keogh and M. J. Pazzani. Derivative Dynamic Time Warping. In *Proceedings of the 2001 SIAM International Conference on Data Mining*, pages 1–11. 2001.