Highway Traffic Flow Measurement by Passive Monitoring of Wi-Fi Signals

Paul Fuxjaeger, Stefan Ruehrup, Hannes Weisgrab

FTW – Telecommunications Research Center Vienna
Donau-City-Str. 1, 1220 Vienna, Austria
{fuxjaeger,ruehrup,weisgrab}@ftw.at

Bernd Rainer

ASFiNAG, Autobahnen- und Schnellstraßen-
Finanzierungs-Aktiengesellschaft, Austria
bernd.rainer@asfinag.at

***Abstract*—Motivated by the fact that a significant number of personal mobile devices are carried into vehicles and that the majority of those devices continuously emit Wi-Fi frames, we investigate the feasibility to use theses transmissions for road traffic analysis. Background transmissions are emitted by the majority of Wi-Fi-enabled devices by means of so called *probe requests*. We show by a real-world measurement on an Austrian motorway that a sufficient number of probe requests can be received in order to re-identify devices traveling between two locations and to estimate travel times. Our results show that by proper post-processing of measurement data, the number of generated travel time estimations per hour are comparable with conventional traffic detectors. We show how to increase user privacy by pruning parts of the Wi-Fi device identifier. This way we avoid a unique relation between a travel time measurement and a Wi-Fi device, which could potentially be linked to a user. Using such ambiguous identifiers has an impact on travel time measurements. Based on our measurements we show a trade-off between privacy and accuracy of the results.**

I. Introduction

Nowadays the penetration of Wi-Fi-enabled personal communication devices such as smartphones and tablets is very high in developed countries. These devices usually transmit control messages in the background that help to discover access points and facilitate connections to other devices. If the those signals can be captured along a motorway and allow to recognize the same device at different locations, it is possible to calculate travel times. Figure 1 illustrates this scenario. Such travel time measurements are an important input to traffic management and travel information systems; they indicate traffic status and allow to give route recommendations.

In general, several methods to identify a vehicle and enable travel time measurements are known: Automatic number plate recognition (ANPR), probe vehicle data (PVD), and radio transmissions. Radio transmissions can be triggered by dedicated transponders, e.g. those used in free-flow tolling systems. We are interested in radio transmissions by smartphones or other consumer devices which are carried in vehicles and equipped with a Wi-Fi interface. The question arises whether enough Wi-Fi devices emitting such signals can be observed on motorways. The prerequisite for the travel time estimation is not only the capability to receive those signals from Wi-Fi devices, to decode them and extract device identifiers—the devices themselves have to emit messages in sufficiently small intervals. To summarize, the feasibility of travel time estimation via monitoring of Wi-Fi signals depends on the following factors:

1) The penetration of Wi-Fi enabled devices in general, 2) the interval between background Wi-Fi transmissions that include unique identifiers and 3) the reception range of the sensor node.

Recording unique devices identifiers at different locations raises the question whether a travel time measurement can be used to reveal mobility information about its user. Basically, there is no direct link between a hardware address of a mobile device and its user. Nevertheless, if we want to make sure that a user cannot be tracked, even if this link exists, the MAC addresses should not be used in its original form. The approach we investigate here is to remove parts of the MAC address in order to make the identifier used in travel time measurements non-unique and its link to a device and its user ambiguous. We will show that accurate travel time measurements are still possible as long as the identifiers remain unique within the sensor network within a given measurement time interval.

The contribution of the paper is as follows:

- We give a detailed technical description of travel time measurements based on re-identification of Wi-Fi devices.
- We provide insights into required data processing steps and compare results to “ground-truth” from a traffic detector.
- We describe a method that helps reducing privacy concerns by using pruned MAC address identifiers and show the impact on travel time measurements.

II. Related Work

Traffic flow and travel time measurements are essential tools in road traffic management. Several approaches exist to detect and count vehicles and measure the traffic flow including loop detectors or radar sensors. Sensors or techniques that can identify individual vehicles such as automatic number-plate recognition (ANPR) or electronic tags allow to measure travel times (see [9] for an overview). Among these approaches, the tracking of vehicles via radio signal emissions is considered a cost-efficient method. Road tolling systems that require the user to carry a transponder in the vehicle are able to deliver travel times at certain measurement points. But also other

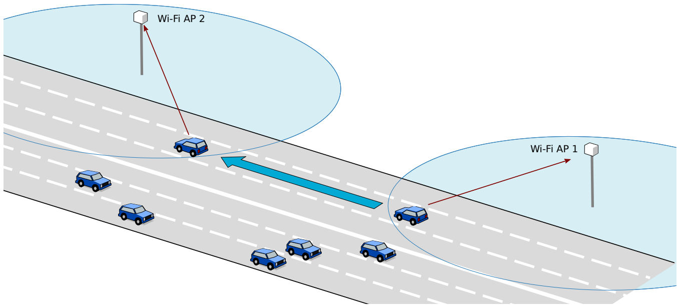

Fig. 1. A basic requirement for travel time estimation is that a Wi-Fi device can be uniquely identified when passing two access points

devices than dedicated transponders emit signals that can be used for road traffic estimation. Most people carry mobile phones which deliver mobility data via the cellular network, however, with low granularity and a need for sophisticated analysis [8]. Smartphones and hands-free car kits emit Bluetooth signals, which can be received by roadside equipment for travel time measurements; BLIP Systems' tracking solution [3] and Swarco's BlueRoadConcept [10] are two commercial examples.

One of the earliest works on Bluetooth and Wi-Fi tracking dates back to 2008 [11] and describes the basic principle, namely the passive monitoring and time-stamping of MAC addresses of wireless devices. The wide-spread use of bluetooth has then led to a focus on Bluetooth tracking [1] and its extensive evaluation in several studies [4], [2], [5]. The recent smartphone trend has put the Wi-Fi technology back on the map for travel time measurement—the basic principle of detecting and time-stamping MAC addresses still remains unchanged.

While the aforementioned studies cover medium and long distance travel times on sub-urban and arterial roads, we put our focus on short-distance travel times under data protection constraints. The data privacy issue has already been raised in the early work by Wasson et al. [11]. Later it has been claimed that MAC addresses need to be encrypted in a one-way fashion at the measurement point, so that subsequent data processing steps work with the encrypted identifier [5]. In the same work it is noted that vendor-specific encryption makes it impossible to calculate travel times across measurement equipment. A standard cryptographic hash function, which is hard to reverse, seems to solve the problem. However, the *allocated part* of the 48bit MAC address space is small and a brute-force approach to reverse an encrypted identifier into the original MAC address is possible [6].

III. Technical Background

A. Wi-Fi Probe Requests

The majority of Wi-Fi networks that comply to the IEEE 802.11 Wireless LAN standard [7] operate in so called infrastructure mode, in which one or more mobile end-user devices are attached to an access point (AP), which in turn is connected to the public Internet.

If an end-user device is not currently attached to an access point, two discovery processes are defined in the standard that are used to detect alternative candidate APs in range:

- **passively** by listening to advertisement messages transmitted by access points. These so called *beacon frames* are sent continuously on the channel the AP is operating on. The default beacon interval configuration setting is 100ms.
- **actively** by transmitting *probe requests* which are answered with probe responses by any AP that overhears the request. The interval of probe request transmission is not standardized.

Every time a Wi-Fi enabled mobile internet device is moved out of the reception range of the AP it is currently attached to, those two mechanisms are triggered to continuously *scan* the environment for alternative connection opportunities.

As of today, the majority of deployed APs are operating in the 2.4 GHz industrial scientific and medical (ISM) spectrum band. The IEEE standard [7] defines 14 distinct channels in this part of the radio frequency spectrum and channel numbers 1 to 11 are allocated for world-wide use by the regulatory bodies. This leads to a severe drawback of the passive scanning method mentioned above: before the attachment process can be started all available Wi-Fi channels need to be exhaustively scanned which takes several seconds to complete.

This discovery time can be reduced by active probing—it is evident that in this case a delicate tradeoff between scanning speedup and reduced battery lifetime (in case of mobile devices) exists.

B. Platform and Device-Specific Behavior

The aforementioned tradeoff (speed versus battery lifetime) and the fact that the interval between probing requests is not standardized suggests that the amount of probe request transmission activity is not uniform over all mobile device platforms. In order to quantify those potential differences we analyzed the two most popular platforms in detail: iOS and Android (which have a combined market share of more than 80% by the end Q4 2013 according to recent market analysis by Gartner Technology Research).

Table I lists all devices that have been tested. The following conclusions can be drawn:

- All devices we analyzed make use of active scanning and generate probe requests when not associated to an access point.
- Probe request transmission behavior is independent of whether a cellular data connection is available or not.
- The interval between probe requests significantly differs amongst both platforms.
- The interval between probe requests also changes from one OS version to the next.
- In case of Android 4.4.2 the probe request transmissions also include the full history of previously used AP identifiers (SSIDs).

In summary, we observed a large diversity amongst tested platforms and OS versions and expect that this heterogeneity

Phone model	Operating System Version	Avg. time between probe requests on channel 1
Google Nexus 7	Android 4.2	300 seconds
Samsung Note 2	Android 4.2	300 seconds
Google Nexus 4	Android 4.4.2	10 seconds
Apple iPhone 5	iOS 6.0	100 seconds
Apple iPhone 5s	iOS 7.1	850 seconds

TABLE I. PROBE REQUEST INTERVALS IN IDLE STATE

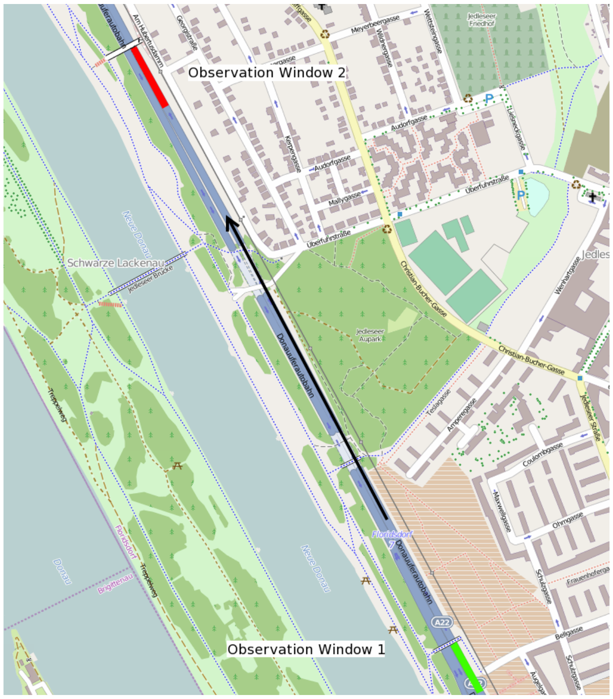

Fig. 2. Map of the two measurement points on highway A22 near Knoten Floridsdorf in the 22nd district of Vienna (©OpenStreetMap contributors).

will keep increasing in the future since the mobile device market is very dynamic and new OS versions and platforms are constantly being released (e.g. Windows Mobile and Firefox OS).

IV. MEASUREMENT SETUP

In order to investigate the feasibility and resulting accuracy of travel-time estimation via Wi-Fi tracking we executed a 30-minute two-point measurement campaign on a highway in the north-east of Vienna.

The central question to be answered was whether enough messages can be received and subsequently correlated over two observation points in order to generate *travel time estimates* for the given highway section.

A. Measurement Location

We choose this section of highway A22 in Vienna because it is covered by pedestrian bridges at multiple points which serve as safe mounting positions while still being very close to the highway lanes. Also, a dedicated ASFiNAG vehicle counting sensor is mounted on a nearby gantry which enabled us to get ground truth data on the total number of vehicles that passed our sensors during measurement time.

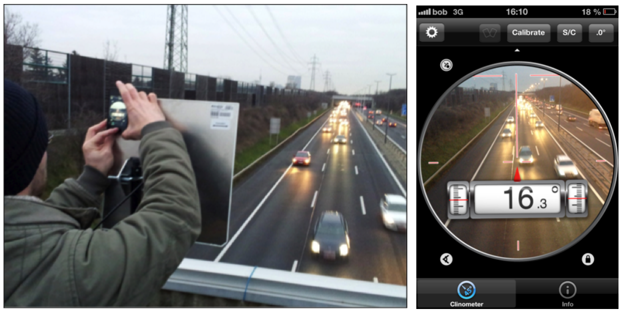

Fig. 3. Setup procedure on the pedestrian bridge at observation point 1.

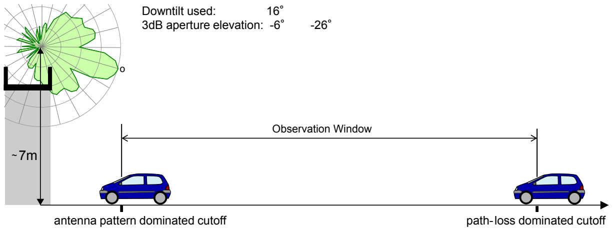
Downtilt used: 16°
3dB aperture elevation: -6° -26°
Observation Window
~7m
antenna pattern dominated cutoff
path-loss dominated cutoff

Fig. 4. Lack of statistics on the total path-loss from the Wi-Fi device through the hull of the vehicles to the receiver antenna makes it impossible to reliably quantify the observation window size.

We positioned two antennas pointing towards vehicles that are driving on the three lanes in north-to-south direction, as depicted in Figure 2 and Figure 3.

B. Sensor Configuration

Evidently, any Wi-Fi signal originating from mobile devices will be significantly dampened by the hull of the vehicle and—if there is no traffic congestion—vehicles themselves are within the reception range of a single sensor unit only for a short period of time (as indicated in Figure 1). This leads to the choice of using high gain directional antennas and aligning them towards to the driving direction, slightly tilted downwards as depicted in Figure 3.

After several re-calibration measurements and subsequent inclination adjustments we settled on a downward tilt of 16 degrees which delivered maximum sensor data output in our case. Unfortunately, the resulting *reception range* is difficult to reliably estimate since many influencing factors are unknown. The main uncertainties are: exact path-loss due to the vehicle hull and exact transmission power, antenna gain and antenna orientation of user devices. This aspect is illustrated in Figure 4.

For both sensor locations we used identical antenna alignments and receiver hardware (see Table II for detailed information), thereby making sure that the observation window length (albeit unknown) is as equal as possible on both locations.

Finally, the recording process was started simultaneously on both locations and the two resulting data sets saved to disk for subsequent (offline) post processing.

V. RESULTS

Table III show the overall results from both observation points; interestingly there is a significant difference between

Highway	A22
Direction	North to South
Date	10th Dec 2013
Speed Limit	80 km/h
Starttime	16:30 h
Stoptime	17:00 h
Distance	1420 m
Total Vehicle Count	1255 (Gantry Sensor Output)
Truck Count	43 (Gantry Sensor Output)
IEEE802.11 Channel	1 (2.412 GHz)
Antenna Model	PAC-24-190
Antenna Gain	19 dBi (at 2.412 GHz)
Polarization	linear
3dB Aperture	20 degrees
Operating System	Ubuntu 12.04 LTS
Receiver model	TP-Link TL-WN722N
Chipset Name	Atheros AR9271
Driver Name	ath9k htc
Chipset Sensitivity	-96 dBm (at 1 Mbit/s)
Receive Mode	802.11b/g/draftn Monitor Mode
Recording Method	tcpdump (libcpap)

TABLE II. MEASUREMENT PARAMETERS

Location		1	2
Total frames captured		8520	4096
Frames containing a MAC SOURCE ADDRESS		4706	3329
thereof	Probe request (0x04)	1696	2620
	Probe response (0x05)	175	174
	Beacon (0x08)	1671	389
	Data (0x20)	263	62
	QoS Data (0x28)	857	77
	Null Function (0x24)	41	7
	De-authentication (0x0c)	3	0

TABLE III. OVERALL MEASUREMENT RESULT STATISTICS

both locations when it comes to the total number of total frames that have been captured. But, as we will demonstrate in the following this difference is merely due to the fact that location 1 was close to a fixed Wi-Fi installation. The frames originating from this non-moving source can be filtered as we will show next.

In the first post-processing stage all frames that do not contain a valid source address in the MAC header are suppressed, which reduces the amount of data by a factor of 0.55 for location 1 and 0.81 for location 2 respectively. In the next step, for those frames that do contain a valid source MAC address we distinguish between two cases: probe requests and all other frame subtypes.

A. Single Point Results

After this classification we aggregate all repetitions of previously overheard source MAC addresses (for each sensor location separately) in order to arrive at a list of unique MAC addresses. The result is outlined in Figure 5 which shows the number of *new unique MAC addresses per minute* that are observed during measurement time at location 1 and location 2 individually. The red line corresponds to the number of new unique identifiers tracked via probe requests whereas the blue line shows the number of new identifiers stemming from all other frame subtypes (carrying source MAC addresses).

Overall, during 30 minute measurement time a total of 428 unique MAC source addresses at location 1 and 531 at location 2 have been recorded via probe request monitoring alone. While only 25 (and 46 respectively) unique addresses are seen due to other Wi-Fi subtype frame transmissions.

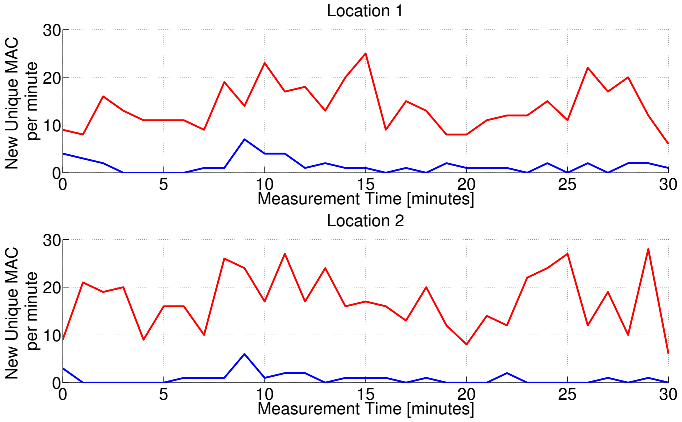

Fig. 5. The average number of new unique MAC source addresses per minute seen at location 1 and location 2 respectively. The red curve corresponds to the case when *only* probe requests are tracked, whereas the blue line is the result for all *other* frame types. The averaging-bin size in these two figures is 10 seconds.

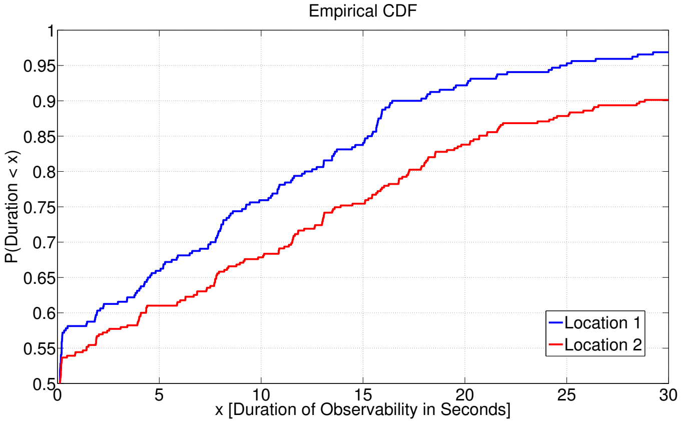

Fig. 6. The empirical cumulative distribution function (ECDF) of observation duration of unique source MAC addresses (on both locations respectively). Change detection of this statistic can be used to detect traffic jams at individual sensor locations. The difference between the two sensor locations in this case is attributed to slightly different observation window sizes (i.e. sensor reception range at location 2 is larger than at location 1).

This result clearly demonstrates that probe request transmissions reveal the majority of unique identifiers and occur at a rate that that may be used fine grained travel time tracking.

Another interesting insight can be be drawn from Figure 6 which plots the empirical cumulative distribution function of the *duration* that single unique MAC addresses are seen at a single sensor location. These statistics change whenever the dwelling time of vehicles inside the reception range of the respective sensor changes. This correlation can be used to detect traffic jams at the location of the sensor—as in this case individual MAC addresses are seen over longer periods of time at the same location.

B. Travel Estimation Results

In the final post-processing stage we jointly analyze both measurement data-sets and evaluate all MAC source addresses that have been observed at *both locations*. In case a match is found we track the time difference and plot the result over measurement time in order to arrive at a continuous stream of

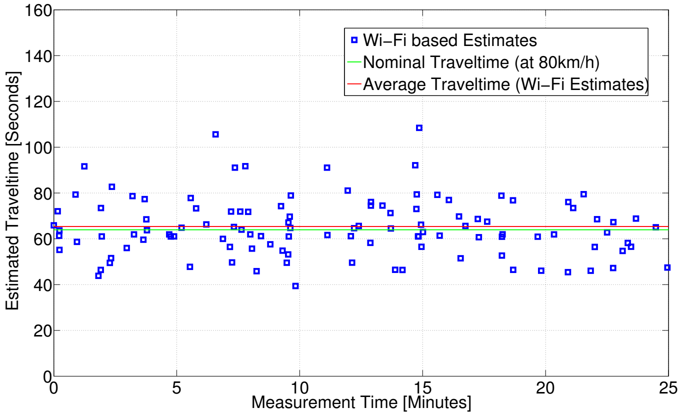

Fig. 7. Final result of the two-point measurement. Based on source MAC addresses correlation (within received probe requests) a total of 138 travel time estimates were generated during 30 minutes observation time.

travel-time estimates. The result is shown in Figure 7 which plots all 138 matches that have been found and displays the average over all estimates during 30 minute measurement time.

It is important to keep in mind that the high variance of the travel-time estimate results in Figure 7 is likely to be caused not only by different individual speeds of vehicles, but also due to the fact that the distance between measurement points is only 1420 m. Since the sensor reception range is unknown and may be in the order of several hundred meters, the exact location of observation cannot be tracked with high sufficient accuracy, which in turn leads to increased relative variance of travel-time results.

During the measurement time a nearby gantry-mounted vehicle counter reported a total of 1255 vehicles passing on the same highway section. Thus, the (unverified) assumption of a simple one-to-one mapping between vehicles and devices leads to the conclusion that travel time estimate data is available for around 11% of all vehicles.

VI. Privacy Aspects

As has been shown in the previous section the correlation of received Wi-Fi probe requests along motorways offers an accurate and efficient method to estimate the overall traffic condition. Using this method, no data-plane information is being processed at all—only MAC addresses within specific Wi-Fi management frames (transmitted in clear text) are used to correlate observations. Furthermore, those identifiers do not reveal any data on personal mobility without additional information on the mapping of MAC addresses to device owners.

Nevertheless, it is evident that privacy-by-design is a desirable architectural feature that should be inherent in any sensor network that is based on data that is generated by individuals. Currently, such design-rules are claimed to be implemented in commercial Wi-Fi tracking systems [3]. The main rationale behind those techniques is that one-way hashing mechanisms can be used to increase the computational effort that is necessary to reverse the mapping between observed MAC addresses and recorded (pseudonymized) data sets in order to extract personal (individual device) mobility traces.

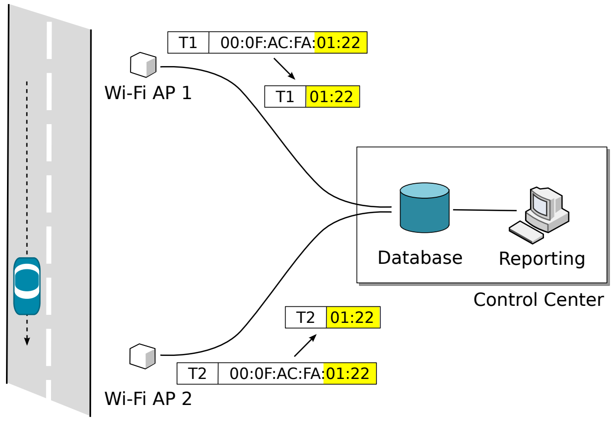

Fig. 8. Two sensors receive frames with a MAC address from a device in a passing vehicle at timestamps T1 and T2. After removing a certain number of bits from the MAC address, both tuples of timestamp and identifier are transmitted to a database, out of which travel time values are extracted.

6 Bytes MAC Address					
0F	CC	2A	00	FE	02
OUI (vendor-specific)			Device ID		

most significant least significant

TABLE IV. MAC Address format

Unfortunately, these claims have already been shown to be untenable since the actually allocated part of the 48bit MAC address space can be reverse-mapped with low computational complexity [6].

In the following we will present and analyze another simple, yet effective method that helps to further decrease the probability that long-term mobility traces of individual Wi-Fi enabled devices can be maliciously extracted from the central data base. Our analysis is based on the following threat model: a malicious party has gained access to (1) the content of the central database via unauthorized access and (2) the mapping of pseudonyms to MAC addresses of devices via exhaustive reverse search [6].

The method we propose is as simple as this: the information that is reported to the central data base is limited to a small fraction of the recorded MAC address. This reduction of data is done directly at the capture points in order to ensure technical unfeasibility of reconstruction of complete MAC addresses at the sensor data fusion center.

The concept is illustrated in Figure 8: identifiers now only contain parts of the MAC address and are not guaranteed to be globally non-unique anymore. Obviously, the loss of global uniqueness leads to a non-zero probability of false-positives if correlation over two sensors is done. And this probability increases as more and more MAC addresses are observed in the respective observation time window. But since the application at hand is highway travel-time estimation a mobility trace becomes irrelevant after a short period of time—the highway operator is only interested in being able to correlate identifiers that have been recorded over a sliding window of several hours.

That means that a Wi-Fi probe-request tracking system that only captures parts of the MAC address may still offer the same benefit for the highway operator and yet does not lead

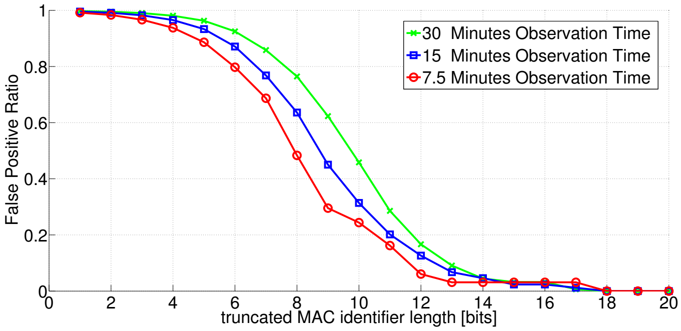

Fig. 9. The ratio of false positives that occur (depending on how many bits of the MAC address are used for matching) for 3 different observation time window lengths.

to a central data base that can be used to extract personal mobility information.

A straightforward method to implement this scheme is to remove the OUI part of the reported MAC address starting (c.f. Table IV). This creates ambiguity because MAC addresses of mobile devices are not globally unique in their last 3 Bytes. This truncation can be even extended to include parts of Device ID field.

After applying this principle on our measurement data set we can report that as few as 18 bits of the device ID part are sufficient in order to retain perfect matching results, the number of false positives (due to potential ambiguity) is still zero in this case. Obviously, since the probability of ambiguity grows as more unique MAC addresses are being captured by sensors over time, 18 bits may not be enough if larger distances (which demand larger sliding observation time windows than 30 minutes) need to be monitored.

In order to quantify the relation between the number of bits necessary for correct detection and the length of the sliding time window the matching is done we compare three different window sizes and plot the resulting false-positive ratio in Figure 9. This result indicates that there exits a linear relation between the length of the sliding observation time window and the required number of bits that are necessary in order to not generate false positives.

Our conjecture is that 20-22 bits (i.e. less than half the available MAC address information) is sufficient for a system that is designed for travel-time tracking in a highway network. The probability of ambiguity (and thus false positives) over a time window of several hours can expected to be negligible in this case as Figure 9 indicates.

VII. Conclusion

Wide-spread use of smart-phones and tablets cause a high amount of Wi-Fi activity on the motorway, which can be used for travel time analysis. A two-point measurement using dedicated equipment was executed to demonstrate feasibility and generate actual travel-time estimates of high accuracy. Our experiments indicate that the current penetration of Wi-Fi enabled devices amongst drivers leads to a probe data output for 11% of all vehicles (assuming a one-to-one relation of devices and vehicles). This method is a very attractive candidate for highway operators to complement their existing set of sensor data sources in order to refine traffic control mechanisms.

We also show that such a sensor system continues to generate valid travel time estimates if only truncated MAC address identifiers are reported to the central data-base. This simple privacy-by-design principle can be used to make extraction of individual long-term mobility traces technically unfeasible.

Although the evolution of background transmission behavior of Wi-Fi enabled smart-phones in the future can not easily predicted, the overall penetration of Wi-Fi enabled devices is expected to continue to grow monotonically for many years to come. Thus, Wi-Fi monitoring is likely to be a suitable option during the next years to increase available travel time data before the large-scale roll-out of cooperative systems based on Car2X communication takes place.

Acknowledgment

This work was performed in the FTW project AMIS, a scientific cooperation co-funded by ASFiNAG. The Austrian Competence Center “FTW Forschungszentrum Telekommunikation Wien GmbH” is funded within the program COMET – Competence Centers for Excellent Technologies by BMVIT, BMWFW, and the City of Vienna. The COMET program is managed by the FFG.

References

[1] B. N. Araghi, L. T. Christensen, R. Krishnan, and H. Lahrmann, “Application of bluetooth technology for mode- specific travel time estimation on arterial roads: Potentials and challenges,” in *Annual Transport Conference at Aalborg University*, ser. ISSN 1603-9696, 2012.

[2] S. R. Aune, “Reisetidsregistrering med blatannteknologi,” Master’s thesis, NTNU - Norwegian University of Science and Technology, 2013.

[3] “http://www.blipsystems.com.”

[4] M. Blogg, C. Semler, M. Hingorani, and R. Troutbeck, “Travel time and origin-destination data collection using bluetooth mac address readers,” in *Australasian Transport Research Forum*, 2010.

[5] S. Cragg, “Bluetooth detection – cheap but challenging,” in *Scottish Transport Applications and Research Conference (STAR)*, Apr. 2013. [Online]. Available: http://www.stsg.org/star/2013/Cragg.pdf

[6] L. Demir, “Wi-fi tracking: what about privacy?” INRIA, Tech. Rep. hal-00859013, version 1, September 2013. [Online]. Available: hal.inria.fr/hal-00859013

[7] “IEEE Std 802.11-2012. part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications,” Mar. 2012.

[8] A. Janecek, K. A. Hummel, D. Valerio, F. Ricciato, and H. Hlavacs, “Cellular data meet vehicular traffic theory: Location area updates and cell transitions for travel time estimation,” in *2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp ’12, 2012.

[9] G. Leduc, “Road traffic data: Collection methods and applications,” European Commission – Joint Research Centre JRC 47967 —- Institute for Prospective Technological Studies., JRC Technical Notes Working Papers on Energy, Transport and Climate Change N.1, 2008.

[10] “Routenqualität online,” DRIVE-ON, p.10-11, http://www.swarco.com/en/content/download/8058/101437/file/DRIVE-ON-1-2011_1.pdf, 2011.

[11] J. Wasson, J. Sturdevant, and D. Bullock, “Real-time travel time estimates using media access control address matching,” *ITE Journal*, June 2008.