Journées thématiques du dpt D2 IRISA

Quantitative Measurement for Location Privacy in IP Vehicular Networks

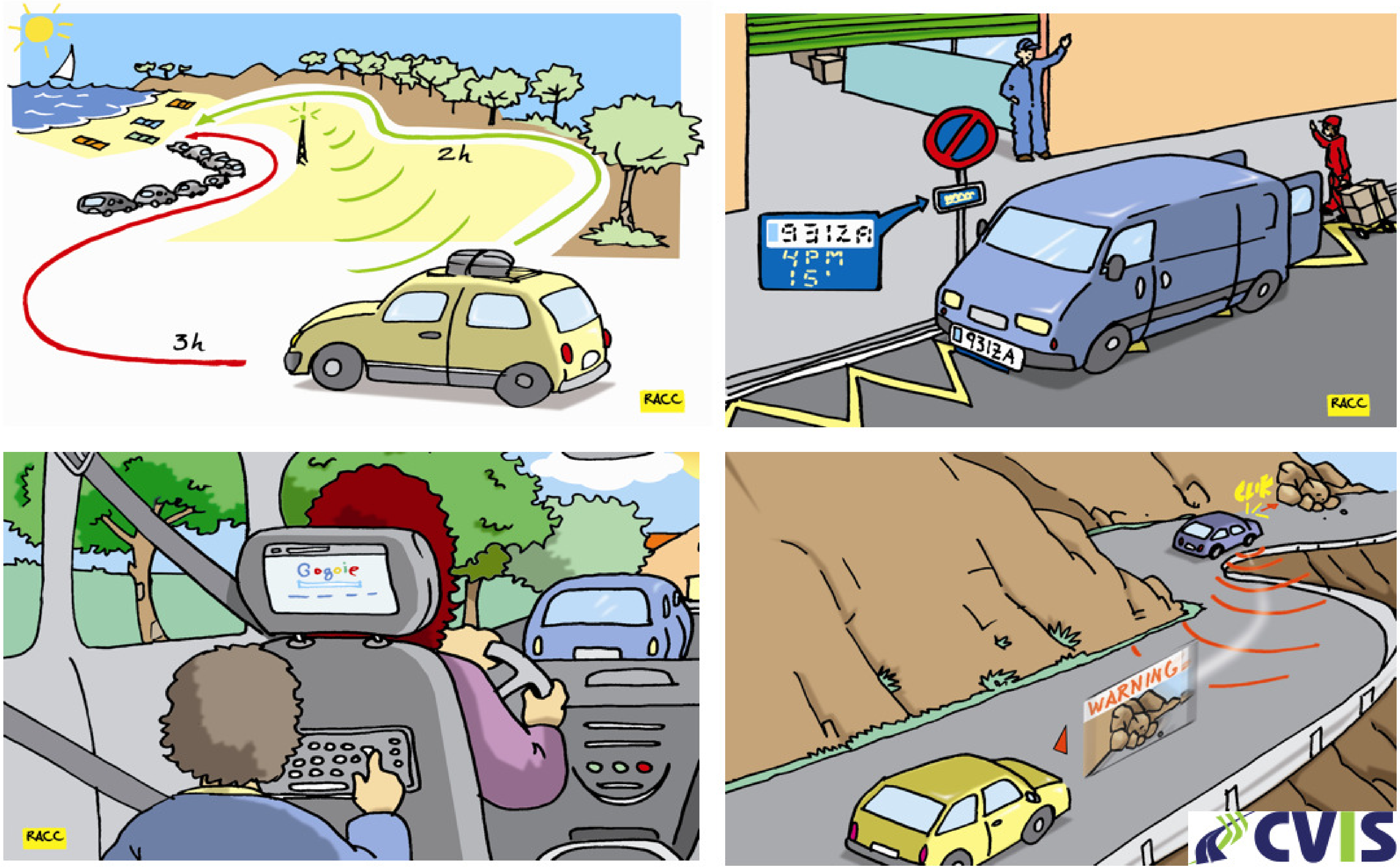
"Optimal value for a pseudonym change"

Jong-Hyouk Lee (jh.lee@telecom-bretagne.eu)

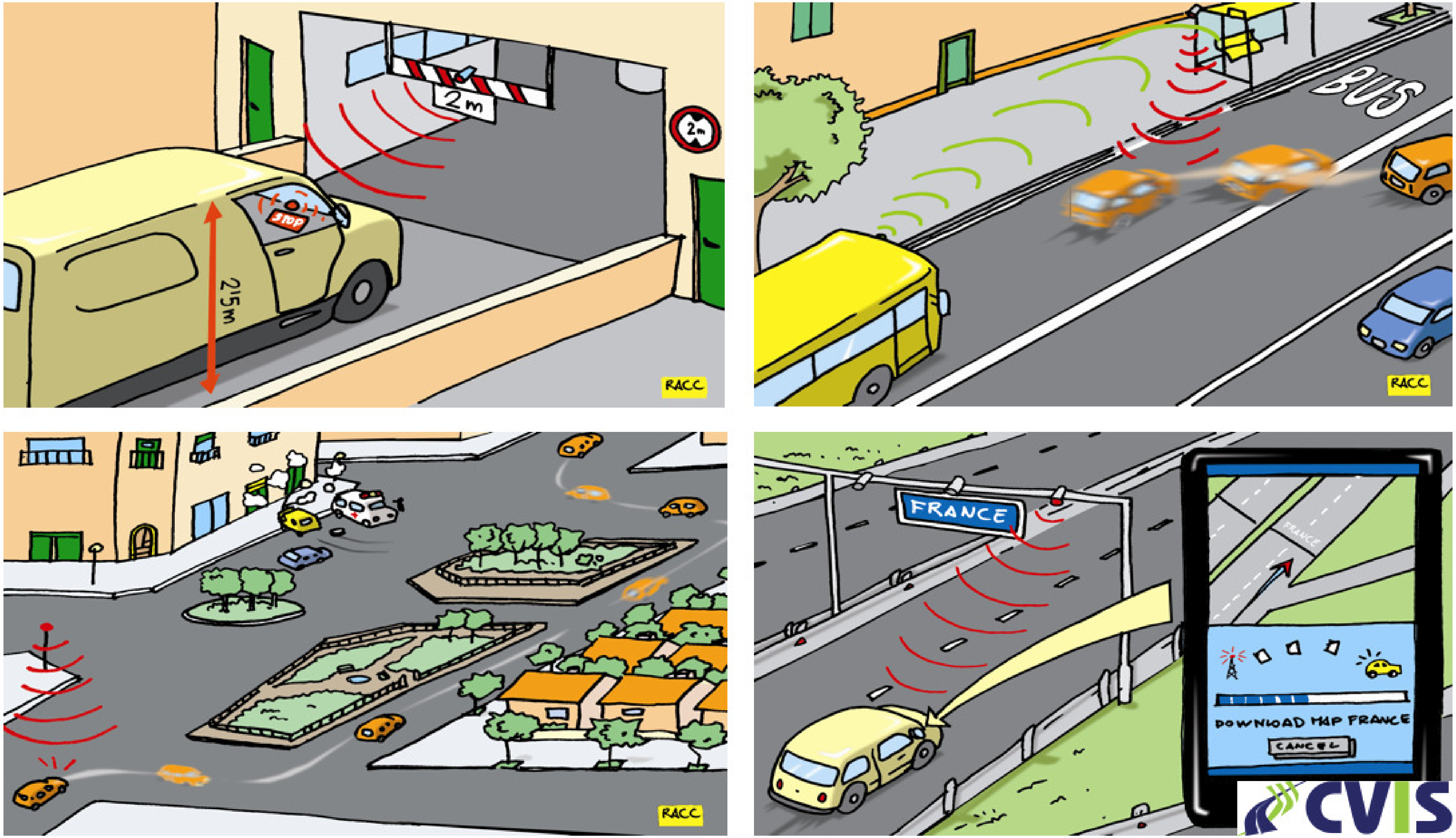
June 28, 2012

- Vehicular Communication (VC)
- Location Privacy Concern
- Pseudonym for Location Privacy
- Pseudonym Change at the IPv6 Layer
- Optimal Value for the Pseudonym Change
- Concluding Remarks

VC providing safety and comfort driving (1/2)

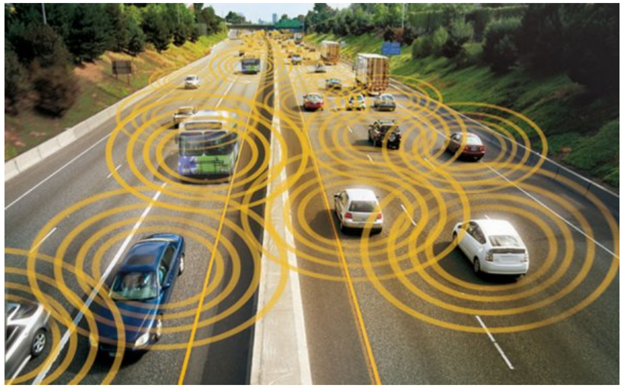


VC providing safety and comfort driving (2/2)



VC is based on wireless communication

VC relies on wireless communication connecting vehicles with roadside infrastructure and with each other



- Communication message contains informative parameters
  - vehicle's location (e.g., GPS info.), heading direction, speed, time, etc
  - message's identification

Message's identification is an address

- For the access (MAC) layer : 48-bit mac address
- For the Geonetworking layer : 64-bit Geonetworking address
- *For the IPv6 layer : 128-bit IPv6 address*

Privacy and Location Privacy

Privacy is a human right to be protected

- 1948 Universal Declaration of Human Rights : Everyone has a right to privacy at home, with family, and in correspondence.

Location privacy is a particular type of information privacy

- Ability to prevent other parties from learning one's current or past location

In VC, as messages contain the identification, location privacy is vulnerable

- Accordingly, *no real identification*, but *temporary identification* in VC
  - based on a set of *pseudonyms*

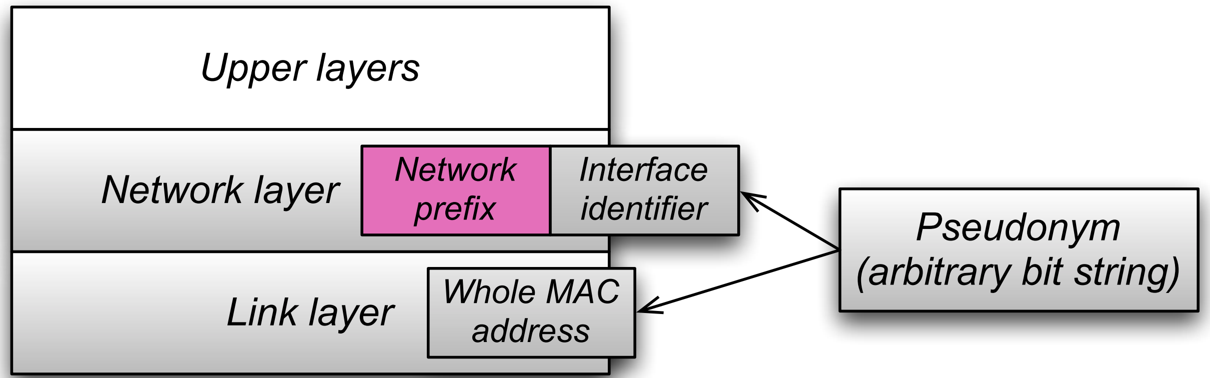
Pseudonym (1/2)

Pseudonym is *an arbitrary bit string*

- to generate *a temporary identifier* for communication

Pseudonym is interpreted in different forms depending on communication layers (protocols) ; in other words, the temporary identifier generated from a pseudonym is used differently

- e.g., *the whole MAC address is replaced* by the temporary identifier while *the interface identifier (rightmost 64-bit) of IPv6 address is replaced* by the temporary identifier



Pseudonym (2/2)

One pseudonym is used only within a short period :

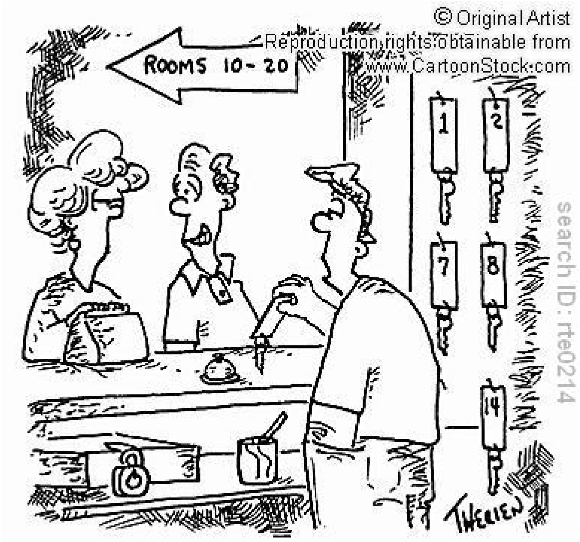
- e.g., a vehicle uses a pseudonym P_i in a short period t_i and changes to a new one P_{i+1} for the next short period t_{i+1} in communication messaging

By using pseudonyms in a short period, attackers (observers) are not able (or at least not easily) to link different messages

- i.e., preventing the attackers from identifying the vehicle emitting messages with pseudonyms

The concept of pseudonym (i.e., use of temporary identifier) is also used in

- 3GPP authentication
- Daily life (!)



"Welcome back, Mr. Smith."

Pseudonym Change at the IPv6 Layer

Pseudonym must be changed :

- Due to *the pseudonym change interval*
  - pseudonym expiration
- Due to *the change of point-of-attachment*
  - vehicle's handover from one access router to another

Pseudonym Change Interval

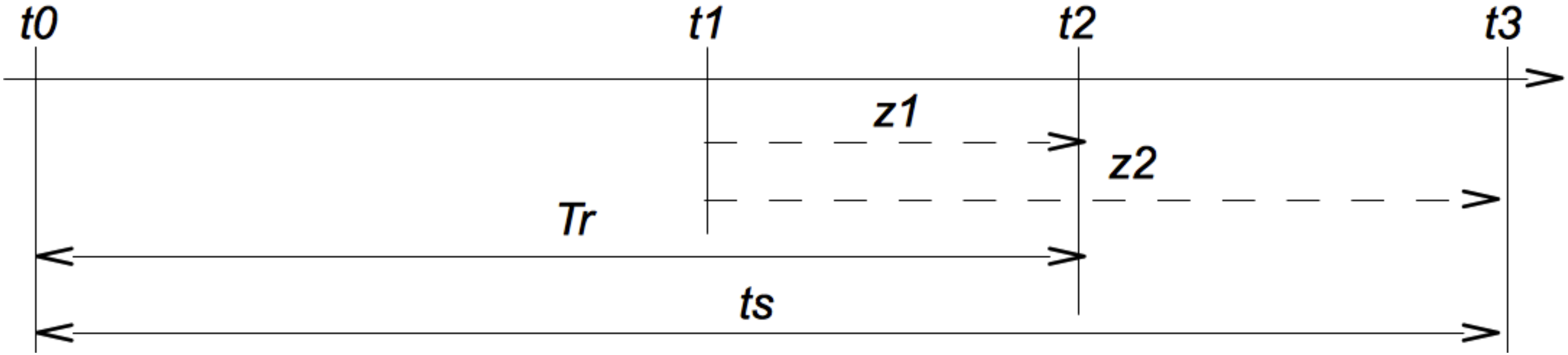
No specific standard for the pseudonym change interval

- *If it is long, the privacy exposed time increases*
- *If it is short, the pseudonym change overhead increases*
  - frequent address generation, validation (i.e., DAD), and registration procedures required
    - communication blocking, packet loss, etc
  - increased network traffic
    - neighbor discovery protocol messages at the access network level
    - binding update/ack messages between the vehicle (mobile router) and home agent (HA)

We need to develop an algorithm that finds *an optimal value for the pseudonym change*

- Making a balance between *performance* and *location privacy* while facilitating VC

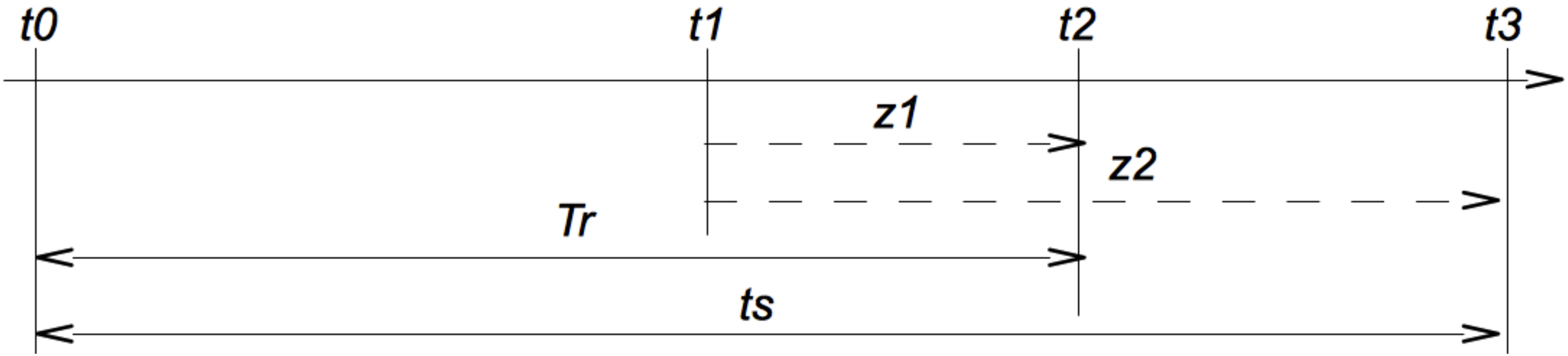
Diagram



- t_0 : the network enter (come in) time
- t_3 : the network leave (come out) time
- $t_s = t_3 - t_0$: the network residence time
- t_1 : observation start time
- t_2 : new pseudonym update time
- T_r : pseudonym change interval

After the observation starts at t_1, the pseudonym is changed (updated) by either at t_2 by a periodical pseudonym change or at t_3 by a handover

Privacy exposed time (1/4)



Suppose z is the privacy exposed time :

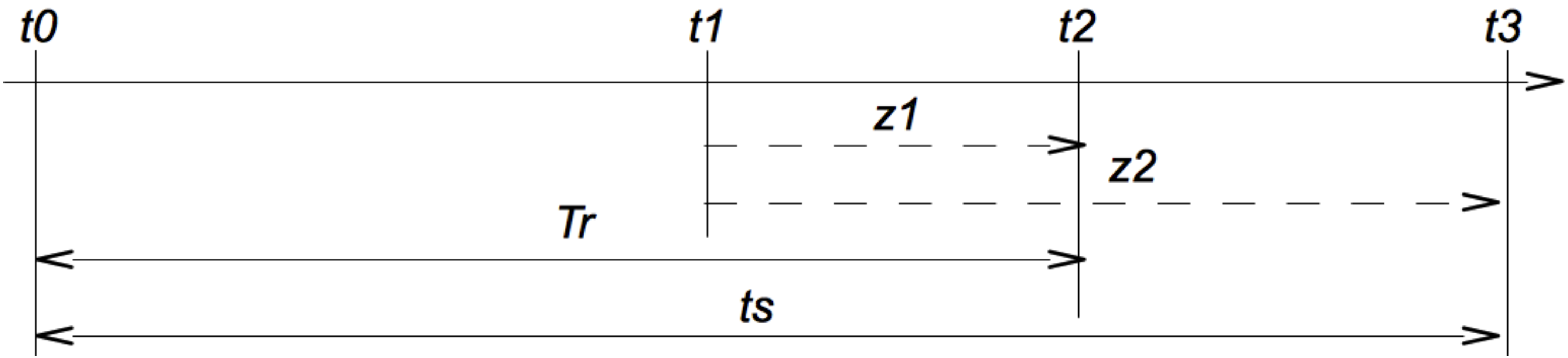
$$z = min\{z_1, z_2\}, \tag{1}$$

where $z_1 = t_2 - t_1 (0 < z_1 < T_r)$ and $z_2 = t_3 - t_1 (0 < z_2 < \infty)$. Then, the PDF of z is given :

$$f(z) = f_1(z) \int_z^{\infty} f_2(t)dt + f_2(z) \int_z^{T_r} f_1(t)dt, \tag{2}$$

where $f_1(z)$ and $f_2(z)$ are PDFs of z_1 and z_2, respectively.

Privacy exposed time (2/4)



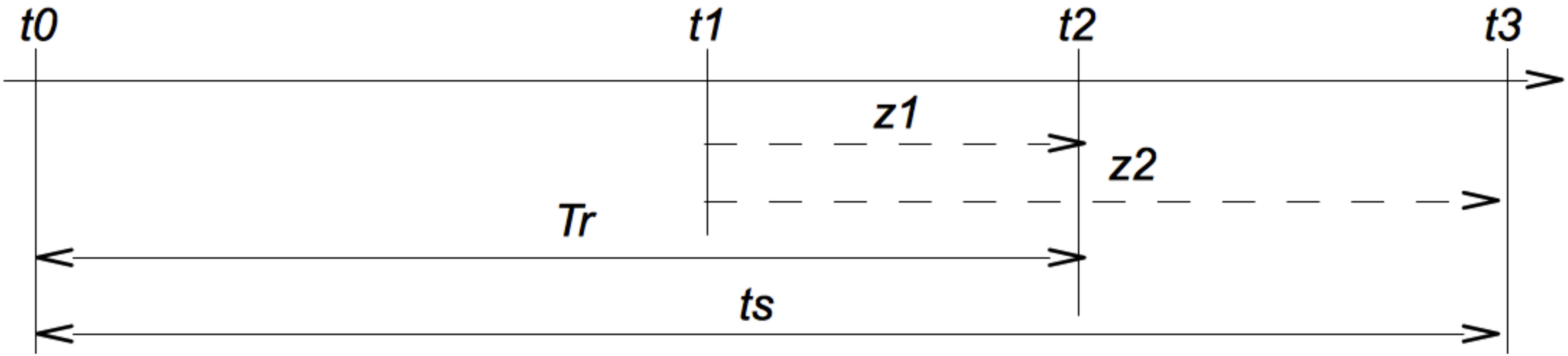
Now, we need to find out proper distributions for z_1 and z_2. In other words, we should have reasonable assumptions for the distributions.

Here, we take a general assumption for z_2 : the network residence time t_s follows an exponential distribution with rate μ_s Then, $f_2(z)$ is calculated as

$$f_2(z) = \mu_s e^{-\mu_s z}. \quad (3)$$

Once we find a distribution that captures realistic behaviors of the observation, i.e., values of z_1, we are ready for obtaining the privacy exposed time in the given model.

Privacy exposed time (3/4)



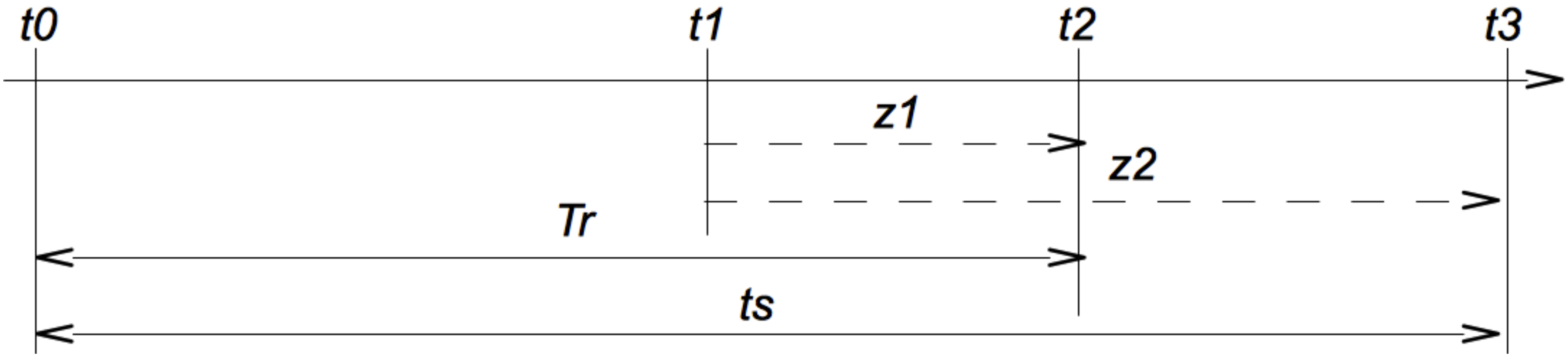
If we assume that z_1 follows a uniform distribution in $[0, T_r]$, $f_1(z)$ is obtained as

$$f_1(z) = \frac{1}{T_r}. \tag{4}$$

Now, we have $f_1(z)$ and $f_2(z)$ that are the PDFs of z_1 and z_2. So, we can obtain the PDF of z (the privacy exposed time), $f(z)$, as

$$\begin{aligned} f(z) &= f_1(z)\int_z^{\infty} f_2(t)dt + f_2(z)\int_z^{T_r} f_1(t)dt \\ &= \frac{1}{T_r}e^{-\mu_s z} + \mu_s e^{-\mu_s z}\frac{1}{T_r}(T_r - z). \end{aligned} \tag{5}$$

Privacy exposed time (4/4)

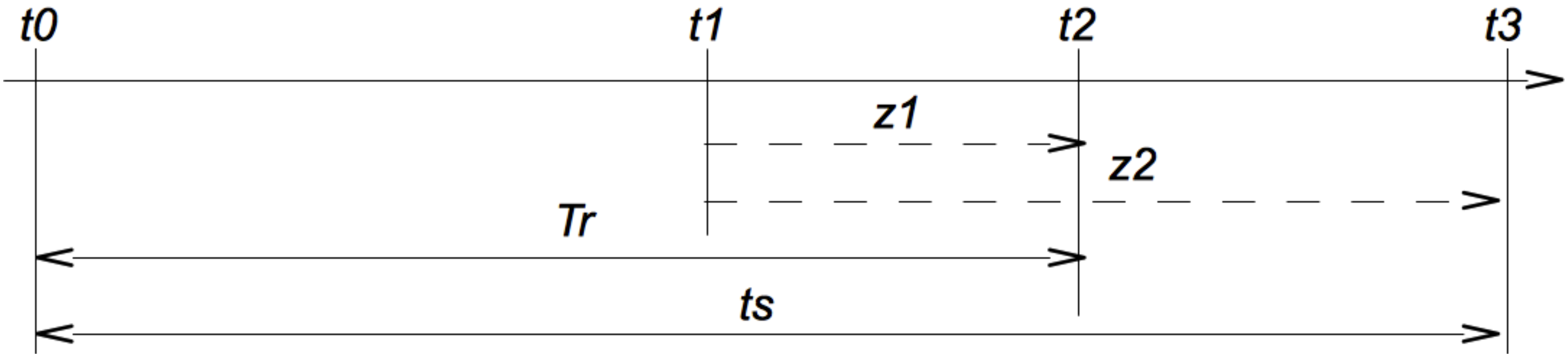


As long as we have the complete form of $f(z)$, we can make the Laplace transform of $f(z)$. Then, by solving the Laplace transform, we will have the expected privacy exposed time, $E[z]$.

Suppose the vehicle generates packets with rate p (packets/sec). Let N denotes the expected number of privacy exposed packets (that use the same pseudonym). Then, it is calculated as

$$N = p \times E[z]. \tag{6}$$

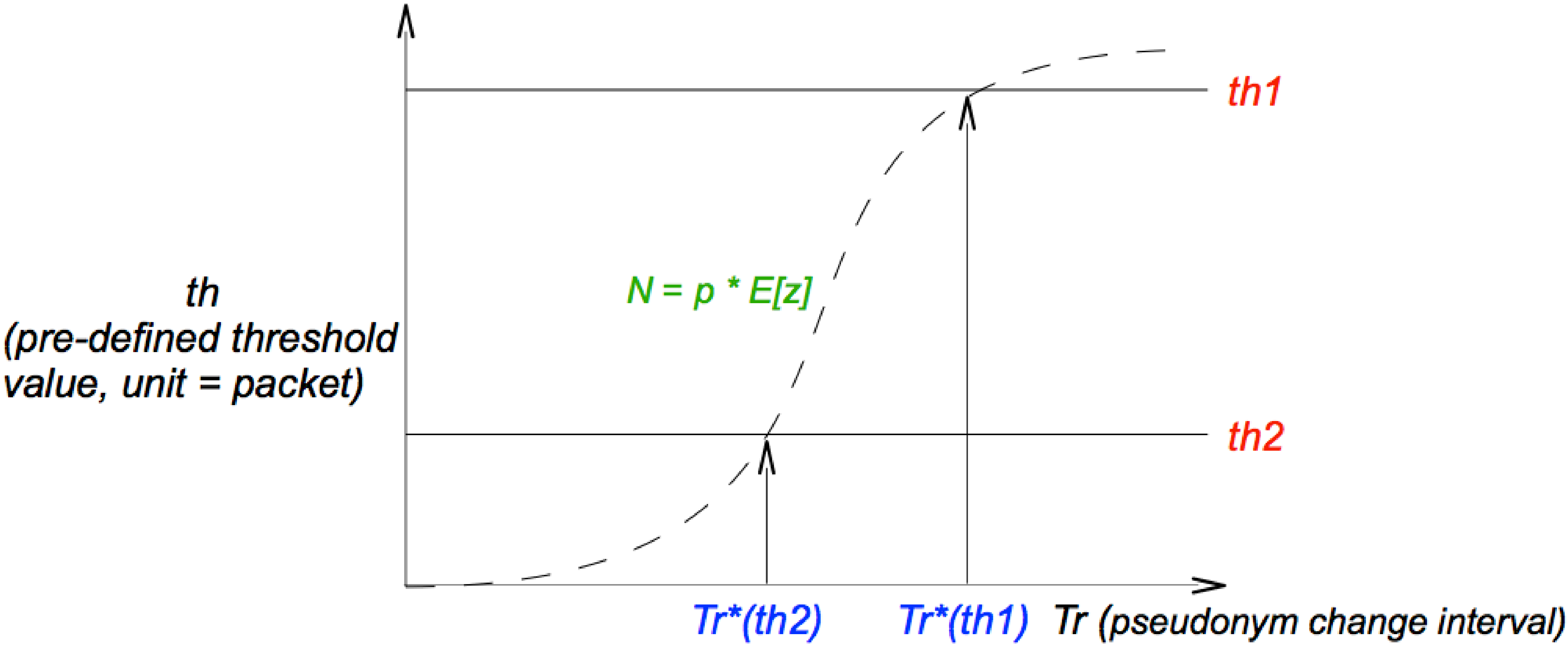
Optimal Pseudonym Change Interval (1/3)



Recall two cases cause the pseudonym change : 1) the pseudonym change interval T_r and 2) the network residence time t_s. Since we cannot control the movement of vehicle, i.e., handover of vehicle, we focus on T_r here to find the optimal interval value for the pseudonym change.

As T_r increases, $N = p \times E[z]$ increases. The optimal value of T_r is thus the maximum value of T_r while N is below a pre-defined threshold value th, which is a security indicator. Note that th should be determined depending on the application types and vehicle's situation (road type, driving time, driving location, etc).

Optimal Pseudonym Change Interval (2/3)



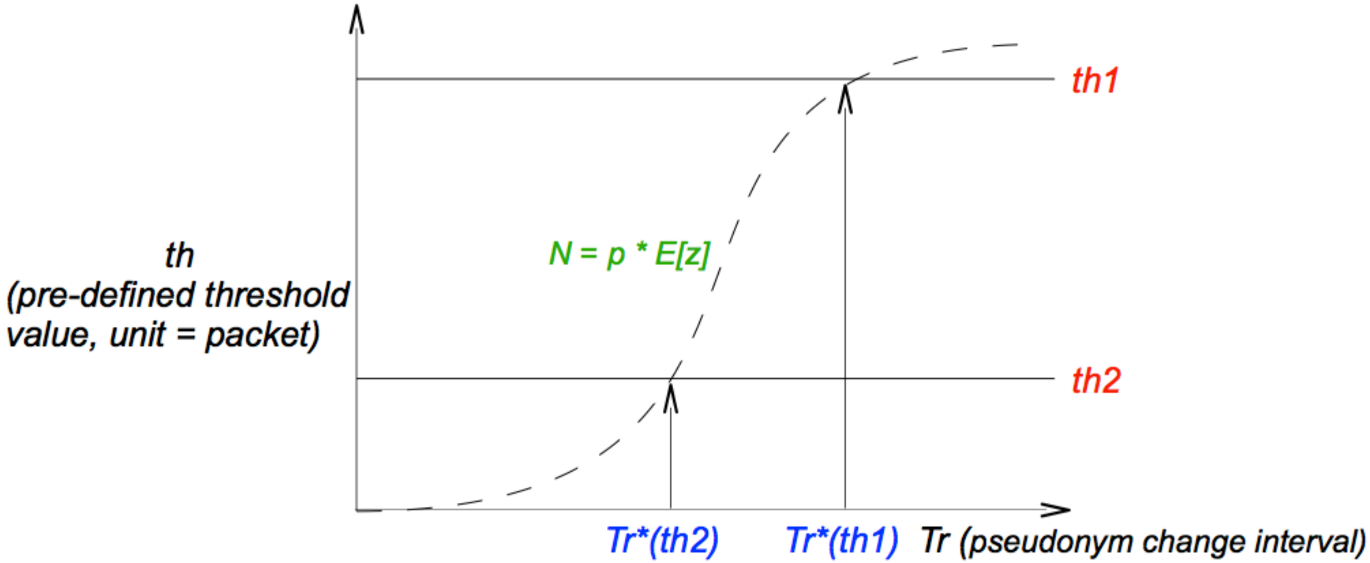
As T_r increases, $N = p \times E[z]$ increases. The optimal value of T_r is thus the maximum value of T_r while N is below a pre-defined threshold value th.

- e.g., $T_r^*(th1)$ is the optimal pseudonym change interval when $th1$ is given, while $T_r^*(th2)$ is the optimal pseudonym change interval with $th2$.

Optimal Pseudonym Change Interval (3/3)

Algorithm 1

$T_r \leftarrow 1, a \leftarrow 0.1;$
Calculate $N \leftarrow p \times E[z];$
while $N < th$ **do**
  $T_r \leftarrow T_r + a;$
  Calculate $N \leftarrow p \times E[z];$
end while
Return $T_r^* \leftarrow (T_r - a);$



Since the above code (algorithm) considers the threshold value *th* (as a security indicator) with the observation behavior $f_1(z)$ and mobility behavior $f_2(z)$, it is seen as an adaptive algorithm to find the optimal pseudonym change interval T_r^*.

We have shown :

- the usage of *pseudonyms* in VC
- the importance of making a balance between *performance* and *location privacy*

An approach for *an optimal pseudonym change* has been briefly introduced :

- with some probability assumptions, e.g., vehicle's residence time
- and limitations, e.g., only IP level pseudonyms are considered

The approach can be further improved :

- with realistic probability assumptions for $f_1(z)$ and $f_2(z)$
- and by studying VC traffic characteristics with security concerns for *th*

Thanks

Thank you for your attention ;)

Jong-Hyouk Lee, *Ph.D, SMIEEE*
jh.lee@telecom-bretagne.eu