APPENDIX

This appendix provides the mathematical proofs of the theoretical results in our paper "**Probabilistic Path Queries in Road Networks: Traffic Uncertainty Aware Path Selection**", in *Proceedings of the Thirteenth International Conference on Extending Database Technology (EDBT'10)*, Lausanne, Switzerland, March 22-26, 2010.

Proof of Theorem 1

PROOF. Since P is a simple path, each edge $e_i \in P$ ($1 \leq i \leq$) is only adjacent to e_{i-1} (if $i > 1$) and e_{i+1} (if $i < n$) in P. Therefore, given w_{e_i}, the weights $w_{e_1}, \ldots, w_{e_{i-1}}$ are conditionally independent on $w_{e_{i+1}}, \ldots, w_{e_n}$. Equation 2 follows with the basic probability theory. ∎

Proof of Theorem 2

PROOF. P_m contains subpath P_{m-2} and edges e_{m-1} and e_m, as illustrated in Figure 11. Therefore,

$$\begin{array}{l} f_{P_m|e_m}(x|y) = Pr[w_{P_{m-1}} = x - y | w_{e_m} = y] \\ = \sum_{z_1+z_2=x-y} Pr[w_{P_{m-2}} = z_1, w_{e_{m-1}} = z_2 | w_{e_m} = y] \end{array}$$

Using the basic probability theory,

$$\begin{array}{l} Pr[w_{P_{m-2}} = z_1, w_{e_{m-1}} = z_2 | w_{e_m} = y] \\ = Pr[w_{e_{m-1}} = z_2 | w_{e_m} = y] \cdot Pr[w_{P_{m-2}} = z_1 | w_{e_{m-1}} = z_2] \end{array}$$

Since $z_1 + z_2 = x - y$, we have

$$\begin{array}{l} Pr[w_{P_{m-2}} = z_1 | w_{e_{m-1}} = z_2] \\ = Pr[w_{P_{m-1}} = x - y | w_{e_{m-1}} = z_2] \end{array}$$

Thus, Equation 5 holds. Equation 6 follows with the basic principles of probability theory. ∎

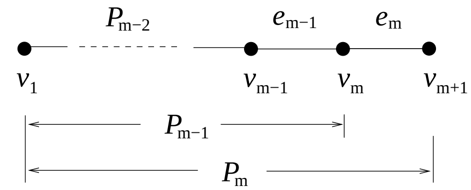
Proof of Lemma 2

PROOF. We prove by contradiction. In the worst case, each bucket contains only one value in w_{P_m}, which means that the probability sum of any two consecutive values in w_{P_m} is greater than $\frac{1}{t}$. Then, if the number of values in w_{P_m} is greater than $2t$, the probability sum of all values in w_{P_m} will be greater than 1, which contradicts the fact that w_{P_m} is a discrete random variable. ∎

Proof of Lemma 3

PROOF. $F'_{P_{m+1}}(l) - F''_{P_{m+1}}(l)$ is the sum of the probabilities in the shaded area in Figures 12(a) and 12(b). In each bucket ϕ_i, the width of the shaded area is $Pr(\phi_i) < \frac{1}{t}$. The sum of the lengthes of all pieces of the shaded area is at most 1. Thus, the probability of the shaded area is at most $\frac{1}{t} \times 1 = \frac{1}{t}$. That is, $F'_{P_{m+1}}(l) - F''_{P_{m+1}}(l) \leq \frac{1}{t}$.

Inequality $|\widehat{F}_{P_{m+1}}(l) - F_{P_{m+1}}(l)| \leq \frac{1}{2t}$ follows immediately. ∎



($w_{P_{m-2}}$ and w_{e_m} are conditionally independent given $w_{e_{m-1}}$)

Figure 11: A path P_m.

Proof of Theorem 3

PROOF. P contains m edges, so $m-1$ steps of bucket approximation are needed to compute the probability distribution of P. We prove the theorem using mathematical induction.

In the first step (computing the probability of $P_3 = \langle v_1, v_2, v_3 \rangle$), we have $|\widehat{F}_{P_3}(l) - F_{P_3}(l)| \leq \frac{1}{2t}$, which is shown in Lemma 3.

Assume that the conclusion holds for the j-th step which computes the probability of P_{j+2} ($j \geq 1$). That is,

$$|\widehat{F}_{P_{j+2}}(l) - F_{P_{j+2}}(l)| \leq \frac{j}{2t} \tag{10}$$

for any real value $l > 0$.

To compute the probability distribution of $w_{P_{j+3}}$, the approximate weight of $w_{P_{j+2}}$ is divided into buckets, such that the probability of each bucket $\phi_i = [x_{z_i}, x_{z'_i}]$ is at most $\frac{1}{t}$. Since the buckets are constructed based on the approximation probability distribution of P_{j+2}, we have

$$\widehat{Pr}(b_i) = \widehat{F}_{P_{j+2}}(x'_i) - \widehat{F}_{P_{j+2}}(x'_{i-1}) \leq \frac{1}{t}.$$

From Inequality 10, we have

$$|\widehat{F}_{P_{j+2}}(x'_i) - F_{P_{j+2}}(x'_i)| \leq \frac{j}{2t}$$

and

$$|\widehat{F}_{P_{j+2}}(x'_{i-1}) - F_{P_{j+2}}(x'_{i-1})| \leq \frac{j}{2t}.$$

The exact probability of ϕ_i is

$$\begin{array}{ll} Pr(\phi_i) & = F_{P_{j+2}|e_{i+2}}(x_{z'_i}|y) - F_{P_{j+2}|e_{i+1}}(x_{z_i}|y) \\ & = \frac{1}{t} + \frac{j}{2t} \times 2 = \frac{j+1}{t}. \end{array}$$

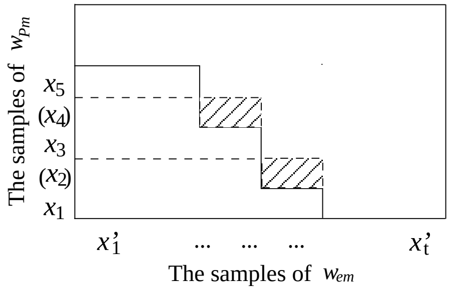
Similar to the proof of Lemma 3, the approximation quality of P_{j+3} in the $(j+1)$-th step can be derived as

$$|\widehat{F}_{P_{j+3}}(x) - F_{P_{j+3}}(x)| \leq \frac{j+1}{2t}.$$

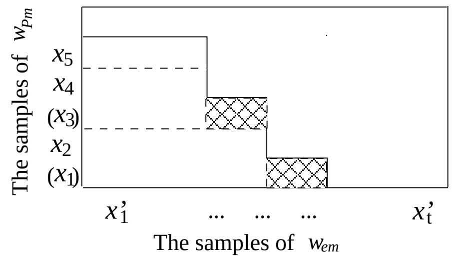
To compute the distribution of P, there are overall $m-1$ steps. Thus, the theorem holds. ∎

Proof of Theorem 4

PROOF. The theorem is an immediate application of the well known Chernoff-Hoeffding bound [1]. ∎



(a) The upper bound.



(b) The lower bound.

Figure 12: The upper/lower bound of $F_{P_{m+1}}(l)$.

Proof of Theorem 5

PROOF. The weight of P is $w_P = w_{P_1} + w_{P_2}$. Therefore,

$$\begin{array}{rl} & F_P(l) = Pr[w_P \leq l] \\ = & Pr[w_{P_1} + w_{P_2} \leq l] \\ = & \sum_{x_1+x_2\leq l} Pr[w_{P_1} = x_1, w_{P_2} = x_2] \\ = & \sum_{x_1+x_2\leq l} Pr[w_{P_1} = x_1] Pr[w_{P_2} = x_2 | w_{P_1} = x_1] \end{array}$$

Since w_{P_1} and w_{P_2} are conditionally independent given w_e, we have

$$Pr[w_{P_2} = x_2 | w_{P_1} = x_1] = \sum_{y \leq x_1} Pr[w_{P_2} = x_2 | w_e = y]$$

Equation 8 follows directly. ■

Proof of Theorem 6

PROOF. (Direction if) If for any path $P = \langle u, \ldots, v_i, \ldots, v \rangle$, $\Delta(v_i, l) \geq F_P(l)$, P* considers all such paths and evaluates their exact $F_P(l)$ values. Therefore, P* can return all paths P such that $F_P(l) \geq \tau$.

(Direction only-if) We prove by contradiction. Assume that there is a path P such that $\Delta(v_i, l) < \tau \leq F_P(l)$. Then, P will not be returned by P* but it is actually an answer path. ■

Proof of Theorem 7

PROOF. Comparing two paths $P = P_1 + P_{opt}$ and $P' = P_1 + P_{2_i}$ ($P_{2_i} \in \mathcal{P}_2$), we have

$$Pr[w_P \leq l] = \sum_{x \leq l} Pr[w_{P_1} = x] \times Pr[w_{P_{opt}} \leq l - x]$$

and

$$Pr[w'_P \leq l] = \sum_{x \leq l} Pr[w_{P_1} = x] \times Pr[w_{P_{2_i}} \leq l - x].$$

Since

$$Pr[w_{P_{opt}} \leq l - x] \geq Pr[w_{P_{2_i}} \leq l - x],$$

we have

$$Pr[w_P \leq l] \geq Pr[w'_P \leq l].$$

■