Supplement to Evaluating Automatic Fault Localization Using Markov Processes

Tim A. D. Henderson,
Google Inc.
1600 Amphitheatre Pkwy
Mountain View, California, USA 94043
tadh@google.com

Yiğit Küçük*, and Andy Podgurski†
Dept. of Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, Ohio, USA 44106
*yigit.kucuk@case.edu †podgurski@case.edu

Abstract

This supplement provides the mathematical proof that motivates the construction of the transition matrices in the main paper. Some terms are restated here for clarity.

I. Restatement of Definitions

The *Standard Rank Score* as defined by Ali *et al.* [1].

Definition 1 (Standard Rank Score)**.** *This score is the expected number of locations a programmer would inspect before locating a fault. Formally, given a set of locations L with their suspiciousness scores $s(l)$ for $l \in L$, the Rank Score for a location $l \in L$ is [1]:*

$$|\{x \ : \ x \in L \wedge s(x) > s(l)\}| + \frac{|\{x \ : \ x \in L \wedge s(x) = s(l)\}|}{2}$$

Note: when we refer to the "Standard Rank Score" this is the definition we are referring to.

A. Background on Ergodic Markov Chains

A finite state Markov chain consists of a set of *states* $S = \{s_1, s_2, ..., s_n\}$ and an $n \times n$ matrix $\mathbf{P}$, called the *transition matrix* [2]. Entry $\mathbf{P}_{i,j}$ gives the probability for a *Markov process* in state s_i to move to state s_j. The probability of a Markov process moving from one state to another only depends on the state the process is currently in. This is known as the *Markov property*.

A Markov chain is said to be *ergodic* if, given enough steps, it can move from any state s_i to any state s_j, i.e. $\Pr\left[s_i \xrightarrow{*} s_j\right] > 0$. Thus, there are no states in an ergodic chain that the process can never leave.

Ergodic Markov chains have *stationary distributions*. Let $\mathbf{v}$ be an arbitrary probability vector. The stationary distribution is a probability vector $\mathbf{w}$ such that

$$\lim_{n \to \infty} \mathbf{v}\mathbf{P}^n = \mathbf{w}$$

The vector $\mathbf{w}$ is a fixed point on $\mathbf{P}$ implying $\mathbf{w}\mathbf{P} = \mathbf{w}$. Stationary distributions give the long term behavior of a Markov chain – meaning that after many steps the chance a Markov process ends in state s_i is given by $\mathbf{w}_i$.

The *expected hitting time* of a state in a Markov chain is the expected number of steps (transitions) a Markov process will make before it encounters the state for the first time. Our new evaluation metric (HT_{Rank}) uses the expected hitting time of the state representing a faulty program location to score a fault localization technique's performance. Lower expected hitting times yield better localization scores.

Definition 2 (Expected Hitting Time)**.** *Consider a Markov chain with transition matrix* $\mathbf{P}$. *Let $T_{i,j}$ be a random variable denoting the time at which a Markov process that starts at state s_i reaches state s_j. The expected hitting time (or just hitting time) of state s_j for such a process is the expected value of $T_{i,j}$*

$$E\left[T_{i,j}\right] = \sum_{k=1}^{\infty} k \cdot Pr\left[T_{i,j} = k\right]$$

B. Expected Hitting Time Rank Score (HT_{Rank})

The new HT_{Rank} score is obtained as follows:

1) A Markov debugging model is supplied (as a Markov chain).
2) The expected hitting times (Def. 2) for each location in the program are computed.
3) The locations are ordered by their expected hitting times.
4) The HT_{Rank} for a location is its position in the ordered list.

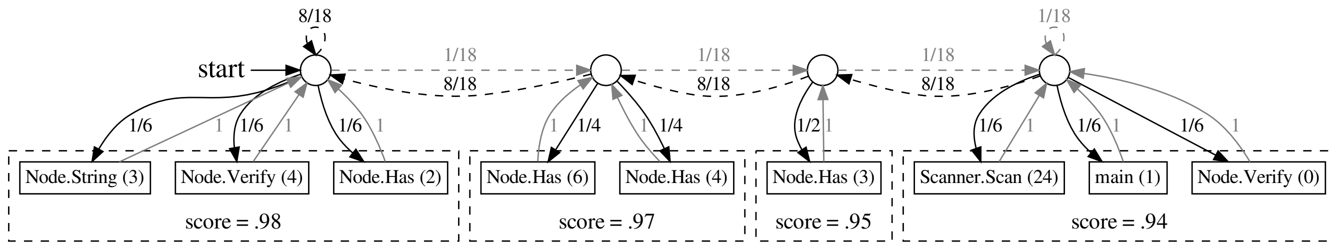

Fig. 1: A simplified version of the Markov model for evaluating the ranked list of suspicious locations for the bug in Listing **??**.

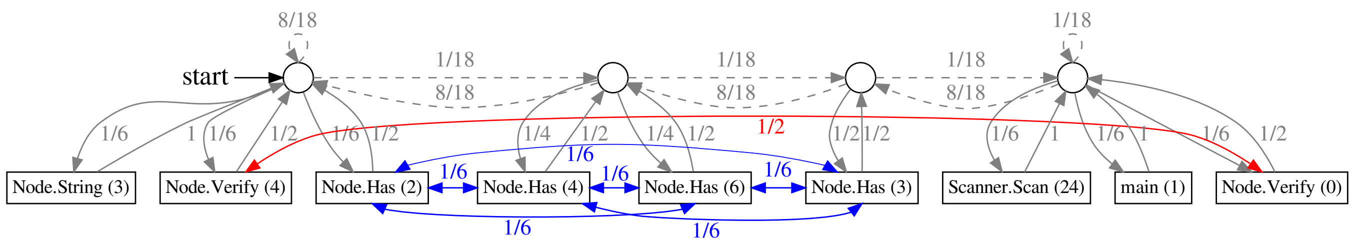

Fig. 2: An example Markov model showing how "jump" edges can be added to represent how a programmer might examine locations which are near the location they are currently reviewing. Compare to Figure 1.

Definition 3 (HT_{Rank})**.** *Given a set of locations L and a Markov chain $(S, \mathbf{P})$ that represents the debugging process and has start state 0, the Hitting-Time Rank Score HT_{Rank} for a location $l \in L \cap S$ is:*

$$\frac{|\{x \ : \ x \in L \cap S \wedge E\left[T_{0,x}\right] < E\left[T_{0,l}\right]\}| + |\{x \ : \ x \in L \cap S \wedge E\left[T_{0,x}\right] = E\left[T_{0,l}\right]\}|}{2}$$

Note: this is almost identical to Definition 1, but it replaces the suspiciousness score with the expected hitting time.

1) A Ranked List as a Markov Chain: Figure 1 provides a graphical example of a Markov chain for a ranked list. Since the nodes in a graphical representation of a Markov chain represent states and the edges represent the transition matrix, the probabilities associated with outgoing edges of each node should sum to 1. In Figure 1, each circular node represents a rank in the list and the square nodes represent associated program locations, which are identified by their function names and static basic-block id number. The square nodes that are grouped together all have the same suspiciousness scores. We will provide here a brief, informal description of the structure of the transition matrix. A formal description of the chain is provided in Definition 4 in the Appendix. The exact choice of transition matrix in the formal chain was driven by a proof of equivalence between HT_{Rank} with this Markov model (as defined in Definition 4) and the Standard Rank Score (Definition 1). The proof is provided below.

The transition matrix boils down to a couple of simple connections. The nodes representing groups form a doubly linked list (see the circular nodes in Figure 1). The ordering of the "group nodes" matches the ordering of the ranks in the ranked list. The links between the node are weighted so that a Markov process will tend to end up in a highly ranked node. More formally, the model was constructed so that if you ordered the Markov states by the probabilities in the *stationary distribution* (which characterizes the long term behavior of the Markov process) from highest to lowest that ordering would match the order of the ranked list.

The second type of connection is from a group node (which once again represents a rank in the ranked lists) to its program location nodes. These nodes once again are Markov states which represent locations in the subject program. Each location node connects to exactly one group node. The transition probabilities (as shown in Figure 1) are set up such that there is an equal chance of moving to any of the locations in the group.

The final connection is from a location node back to its group node. This is always assigned probability 1 (see Figure 1). Again, see Definition 4 in the Appendix for the formal description.

Definition 4 (CBSFL Ranked List Markov Chain)**.** *To construct a Markov chain representing a list of program locations ranked in descending order by their suspiciousness scores:*

1) *Let L be the set of locations in the program.*
2) *For a location $l \in L$ let $s(l)$ be its CBSFL suspiciousness score.*

3) *Partition the locations in L into a list of groups $G = \{g_1 \subseteq L, g_2 \subseteq L, ..., g_n \subseteq L\}$ such that for each group g_i all of the locations it contains have the same score: $\forall\ g_i \in G,\ \forall\ l, l' \in g_i\ [s(l) = s(l')]$*
4) *The score of a group $s(g_i)$ is defined to be the common score of its members: $\forall\ l \in g_i\ [s(g_i) = s(l)]$*
5) *Order G by the scores of its groups, such that g_0 has the highest score and g_n has the lowest: $s(g_0) > s(g_1) > ... > s(g_n)$*
6) *Now construct the set of states. There is one state for each group $g \in G$ and for each location $l \in L$.*

$$S = \{g : g \in G\} \cup \{l : l \in L\}$$

7) *Finally construct the transition matrix $\mathbf{P}$ for the states S.*

$$\mathbf{P}_{i,j} = \begin{cases} 1 & \text{if} \quad s_i \in L \wedge s_j \in G \wedge s_i \in s_j \\ \frac{|L|-1}{2|L|} & \text{if} \quad s_i = g_0 \wedge s_j = s_i \\ \frac{1}{2|L|} & \text{if} \quad s_i = g_n \wedge s_j = s_i \\ \frac{|L|-1}{2|L|} & \text{if} \quad s_i \in G \wedge s_j \in G \wedge s_i - 1 = s_j \\ \frac{1}{2|L|} & \text{if} \quad s_i \in G \wedge s_j \in G \wedge s_i + 1 = s_j \\ \frac{1}{2|s_i|} & \text{if} \quad s_i \in G \wedge s_j \in L \wedge s_j \in s_i \\ 0 & \text{otherwise} \end{cases}$$

II. Proof of: CBSFL RankScore = HT_{RANK} for the chain in Defintion 4

Now, it will be shown that for the chain in Definition 4, which represents the assumed CBSFL debugging process, and for all locations $l \in L$ the ordinary CBSFL Rank Score $\text{RankScore}(l)$ is the same as new $\text{HT}_{\text{Rank}}(L, S, \mathbf{P}, l)$ score. This proof is important because it establishes that the Markov chain in Definition 4 correctly models the rank list debugging scenario. As will be seen in the proof of Lemma 2 below, the choice of transition probabilities made during the construction of the chain was important to ensure that the asymptotic behavior of the chain converges toward visiting locations appearing earlier in the rank list before visiting locations that appear later. The proof consists of 4 supporting lemmas (in Appendix **??** due to their length) and culminates in Theorem 1 which asserts that CBSFL Rank Score and Hitting Time Rank Score are equivalent.

Theorem 1. *For a location $l \in L$ with a suspicious score of $s(l)$ the CBSFL Rank Score corresponds with the Hitting-Time Rank Score for the Chain defined in Definition 4.*

Proof. This is a sketch. The details are given in 4 lemmas in Appendix II. To show $\text{HT}_{\text{Rank}} = \text{RankScore}$ begin by writing:

$$\begin{aligned} &\left|\left\{x \ : \ x \in S \wedge x \in L \wedge t_{\text{hit}[\mathbf{P}]}(0,x) < t_{\text{hit}[\mathbf{P}]}(0,l)\right\}\right| \\ &+ \frac{\left|\left\{x \ : \ x \in S \wedge x \in L \wedge t_{\text{hit}[\mathbf{P}]}(0,x) = t_{\text{hit}[\mathbf{P}]}(0,l)\right\}\right|}{2} \\ = \ &|\{x \ : \ x \in L \wedge s(x) > s(l)\}| + \frac{|\{x \ : \ x \in L \wedge s(x) = s(l)\}|}{2} \end{aligned}$$

The statement is true when the statement

$$\begin{aligned} &\left|\left\{x \ : \ x \in S \wedge x \in L \wedge t_{\text{hit}[\mathbf{P}]}(0,x) < t_{\text{hit}[\mathbf{P}]}(0,l)\right\}\right| \\ &= |\{x \ : \ x \in L \wedge s(x) > s(l)\}| \end{aligned}$$

is true and the statement

$$\begin{aligned} &\frac{\left|\left\{x \ : \ x \in S \wedge x \in L \wedge t_{\text{hit}[\mathbf{P}]}(0,x) = t_{\text{hit}[\mathbf{P}]}(0,l)\right\}\right|}{2} \\ &= \frac{|\{x \ : \ x \in L \wedge s(x) = s(l)\}|}{2} \end{aligned}$$

is also true. The statements are true if for all locations a and b

$$\begin{aligned} t_{\text{hit}[\mathbf{P}]}(0,a) = t_{\text{hit}[\mathbf{P}]}(0,b) &\iff s(a) = s(b) \\ t_{\text{hit}[\mathbf{P}]}(0,a) < t_{\text{hit}[\mathbf{P}]}(0,b) &\iff s(a) > s(b) \end{aligned}$$

We prove these statements. The first statement is shown in Lemmas 1 and 4. The second statement is shown in Lemmas 2 and 3. Therefore, the Hitting Time Rank Score for the chain in Definition 4 is equivalent to the CBSFL RankScore.

□

Lemma 1. *For the chain in Definition 4 if locations $x, y \in L$ have the same suspiciousness score $s(x) = s(y)$ then:*

$$t_{hit[\mathbf{P}]}(0, x) = t_{hit[\mathbf{P}]}(0, y)$$

Proof. First note that by construction of the chain in Definition 4 the locations x and y will be attached to the same group state g_i (see Figure 1). By construction, there is only a process starting at g_i can transition to x or y in one step:

$$\forall\ s \in S \left[s \neq g_i \implies \Pr\left[s \xrightarrow{1} \{x, y\} \right] = 0 \right]$$

$$\Pr\left[g_i \xrightarrow{1} \{x, y\} \right] > 0$$

By construction:

$$\Pr\left[g_i \xrightarrow{1} x \right] = \Pr\left[g_i \xrightarrow{1} y \right]$$

From the definition of $t_{\text{hit}[\mathbf{P}]}(i, j)$ and Theorem **??** we get:

$$t_{\text{hit}[\mathbf{P}]}(0, j) = \sum_{s_k \in T} \mathrm{E}\left[Y_{0,k}^{(0)} \right] + \mathrm{E}\left[Y_{0,k}^{(1)} \right] + \mathrm{E}\left[Y_{0,k}^{(2)} \right] + ...$$

Using the expectation definition of hitting time prove

$$\sum_{k \neq x} \mathrm{E}\left[Y_{0,k}^{(0)} \right] + \mathrm{E}\left[Y_{0,k}^{(1)} \right] + \mathrm{E}\left[Y_{0,k}^{(2)} \right] + ... = \sum_{k \neq y} \mathrm{E}\left[Y_{0,k}^{(0)} \right] + \mathrm{E}\left[Y_{0,k}^{(1)} \right] + \mathrm{E}\left[Y_{0,k}^{(2)} \right] + ...$$

For all $k \neq x$ or y, the terms of the summations are identical and cancel out leaving:

$$\mathrm{E}\left[Y_{0,y}^{(0)} \right] + \mathrm{E}\left[Y_{0,y}^{(1)} \right] + \mathrm{E}\left[Y_{0,y}^{(2)} \right] + ... = \mathrm{E}\left[Y_{0,x}^{(0)} \right] + \mathrm{E}\left[Y_{0,x}^{(1)} \right] + \mathrm{E}\left[Y_{0,x}^{(2)} \right] + ...$$

Which gives a sequence of statements to prove

$$\begin{aligned} \mathrm{E}\left[Y_{0,y}^{(0)} \right] &= \mathrm{E}\left[Y_{0,x}^{(0)} \right] \\ \mathrm{E}\left[Y_{0,y}^{(1)} \right] &= \mathrm{E}\left[Y_{0,x}^{(1)} \right] \\ \mathrm{E}\left[Y_{0,y}^{(2)} \right] &= \mathrm{E}\left[Y_{0,x}^{(2)} \right] \\ &... \end{aligned}$$

These can be restated using the definition for the expectation as:

$$\begin{aligned} \left[{}^{\alpha(x)}\mathbf{P} \right]_{0,y}^{0} &= \left[{}^{\alpha(y)}\mathbf{P} \right]_{0,x}^{0} \\ \left[{}^{\alpha(x)}\mathbf{P} \right]_{0,y}^{1} &= \left[{}^{\alpha(y)}\mathbf{P} \right]_{0,x}^{1} \\ \left[{}^{\alpha(x)}\mathbf{P} \right]_{0,y}^{2} &= \left[{}^{\alpha(y)}\mathbf{P} \right]_{0,x}^{2} \\ &... \end{aligned}$$

Since, $\mathbf{P}^0 = \mathbf{I}$ and $0 \neq x$ or y the statement 0 is true.

$$\left[{}^{\alpha(x)}\mathbf{P} \right]_{0,y}^{0} = \left[{}^{\alpha(y)}\mathbf{P} \right]_{0,x}^{0} \quad \checkmark$$

By construction

$$\forall_k \left(\left[{}^{\alpha(x)}\mathbf{P} \right]_{k,y} = \mathbf{P}_{k,y} = \left[{}^{\alpha(y)}\mathbf{P} \right]_{k,x} = \mathbf{P}_{k,x} \right)$$

Thus statement 1 is true

$$\left[{}^{\alpha(x)}\mathbf{P} \right]_{0,y} = \left[{}^{\alpha(y)}\mathbf{P} \right]_{0,x} \quad \checkmark$$

For every other statement i for $i = 2...$ write

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i}_{0,y} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i}_{0,x}$$

$$\sum_k \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,k} \left[{}^{\alpha(x)}\mathbf{P}\right]_{k,y} = \sum_k \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,k} \left[{}^{\alpha(y)}\mathbf{P}\right]_{k,x}$$

By the construction of the chain

$$\forall\ k \neq g_i \left[\left[{}^{\alpha(x)}\mathbf{P}\right]_{k,y} = \left[{}^{\alpha(y)}\mathbf{P}\right]_{k,x} = 0\right]$$

Yielding

$$\sum_k \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,k} \left[{}^{\alpha(x)}\mathbf{P}\right]_{k,y} = \sum_k \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,k} \left[{}^{\alpha(y)}\mathbf{P}\right]_{k,x}$$

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_i} \left[{}^{\alpha(x)}\mathbf{P}\right]_{g_i,y} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_i} \left[{}^{\alpha(y)}\mathbf{P}\right]_{g_i,x}$$

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_i} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_i}$$

To prove the last expression we will prove the more general version for any group g

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g}$$

With the inductive hypothesis for any group g the statement for $i-1$ is true. Now for the first base case $i = 0$

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{0}_{0,g} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{0}_{0,g}$$

$$\mathbf{I}_{0,g} = \mathbf{I}_{0,g} \quad \checkmark$$

For case $i = 1$

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{1}_{0,g} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{1}_{0,g}$$

$$\mathbf{P}_{0,g} = \mathbf{P}_{0,g} \quad \checkmark$$

For the inductive case $i > 1$

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i}_{0,g} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i}_{0,g}$$

$$\sum_k \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,k} \left[{}^{\alpha(x)}\mathbf{P}\right]_{k,g} = \sum_k \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,k} \left[{}^{\alpha(y)}\mathbf{P}\right]_{k,g}$$

Using the chain's construction the sum has 3 parts for both $\alpha(x)$ and $\alpha(y)$. The first part is for the transition for $(g-1)$, second for $g+1$, and the rest for all of the locations in the group represented by the state g. Demonstrated for the $\alpha(x)$ side (but symmetrical for $\alpha(y)$).

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,(g-1)} \left[{}^{\alpha(x)}\mathbf{P}\right]_{(g-1),g} + \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,(g+1)} \left[{}^{\alpha(x)}\mathbf{P}\right]_{(g+1),g} + \sum_{l \in g} \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l} \left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g}$$

Dealing with each term of the expression individually we have

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_{(i-1)}} \left[{}^{\alpha(x)}\mathbf{P}\right]_{g_{(i-1)},g_i} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_{(i-1)}} \left[{}^{\alpha(y)}\mathbf{P}\right]_{g_{(i-1)},g_i}$$

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_{(i+1)}} \left[{}^{\alpha(x)}\mathbf{P}\right]_{g_{(i+1)},g_i} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_{(i+1)}} \left[{}^{\alpha(y)}\mathbf{P}\right]_{g_{(i+1)},g_i}$$

$$\sum_{l \in g_i} \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l} \left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g_i} = \sum_{l \in g_i} \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,l} \left[{}^{\alpha(y)}\mathbf{P}\right]_{l,g_i}$$

For the first two statements above note the construction of the absorbing chain does not change the probability of moving from a state representing one group to another state representing a group.

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_{(i-1)}} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_{(i-1)}}$$
$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_{(i+1)}} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_{(i+1)}}$$

Using the inductive hypothesis those statements are true. This leaves the statement

$$\sum_{l\in g}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g} = \sum_{l\in g}\left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g}$$

Note, if g is equal to g_i then by the construction of the absorbing chain $\alpha(x)$ (and symmetrical for $\alpha(y)$) $\left[{}^{\alpha(x)}\mathbf{P}\right]_{x,g_i} = 0$

$$\sum_{l\in g_i}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g_i} = \sum_{l\in g_i\wedge l\neq x}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g_i}$$

By the construction of the chain the rest of the probabilities for returning to g_i are 1, $\left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g_i} = 1$

$$\sum_{l\in g_i\wedge l\neq x}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{P}\right]_{l,g_i} = \sum_{l\in g_i\wedge l\neq x}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l}$$

Rewriting

$$\begin{aligned}\sum_{l\in g_i\wedge l\neq x}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l} &= \sum_{l\in g_i\wedge l\neq x}\left(\sum k\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,k}\left[{}^{\alpha(x)}\mathbf{P}\right]_{k,l}\right)\\ &= \sum_{l\in g_i\wedge l\neq x}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,g_i}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g_i,l}\\ &= (|g_i|-1)\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,g_i}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g_i,y}\end{aligned}$$

If g is not equal to g_i we have (where the last l is any $l \in g$).

$$\begin{aligned}\sum_{l\in g}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,l} &= \sum_{l\in g}\left(\sum k\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,k}\left[{}^{\alpha(x)}\mathbf{P}\right]_{k,l}\right)\\ &= \sum_{l\in g}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,g}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g,l}\\ &= (|g|)\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,g}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g,l}\end{aligned}$$

So the expressions for $g \neq g_i$ and g_i are (where l is any location in g).

$$(|g_i|-1)\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,g_i}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g_i,y} = (|g_i|-1)\left[{}^{\alpha(y)}\mathbf{P}\right]^{i-2}_{0,g_i}\left[{}^{\alpha(y)}\mathbf{P}\right]_{g_i,x}$$
$$(|g|)\left[{}^{\alpha(x)}\mathbf{P}\right]^{i-2}_{0,g}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g,l} = (|g|)\left[{}^{\alpha(y)}\mathbf{P}\right]^{i-2}_{0,g}\left[{}^{\alpha(y)}\mathbf{P}\right]_{g,l}$$

Using the inductive hypothesis again

$$\left[{}^{\alpha(x)}\mathbf{P}\right]_{g_i,y} = \left[{}^{\alpha(y)}\mathbf{P}\right]_{g_i,x}$$
$$\left[{}^{\alpha(x)}\mathbf{P}\right]_{g,l} = \left[{}^{\alpha(y)}\mathbf{P}\right]_{g,l}$$

Those statements are true by the construction of the chain and the absorbing chains. This proves the induction for any group g

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{i}_{0,g} = \left[{}^{\alpha(y)}\mathbf{P}\right]^{i}_{0,g} \quad \checkmark$$

Which in turn proves

$$\begin{aligned}\left[{}^{\alpha(x)}\mathbf{P}\right]^{i}_{0,y} &= \left[{}^{\alpha(y)}\mathbf{P}\right]^{i}_{0,x}\\ \sum_k \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,k}\left[{}^{\alpha(x)}\mathbf{P}\right]_{k,y} &= \sum_k \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,k}\left[{}^{\alpha(y)}\mathbf{P}\right]_{k,x}\\ \left[{}^{\alpha(x)}\mathbf{P}\right]^{i-1}_{0,g_i}\left[{}^{\alpha(x)}\mathbf{P}\right]_{g_i,y} &= \left[{}^{\alpha(y)}\mathbf{P}\right]^{i-1}_{0,g_i}\left[{}^{\alpha(y)}\mathbf{P}\right]_{g_i,x} \quad \checkmark\end{aligned}$$

Which completes the proof of the lemma

$$s(x) = s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) = t_{\text{hit}[\mathbf{P}]}(0,y)$$

□

Lemma 2. *For the chain in Definition 4 if location $x \in L$ has a larger suspiciousness score than $y \in L$, $s(x) > s(y)$, then:*

$$t_{\textit{hit}[\mathbf{P}]}(0,x) < t_{\textit{hit}[\mathbf{P}]}(0,y)$$

Proof. Note, from Lemma 4 since $s(x) \neq s(y)$ the hitting times for x and y are different.

The hitting time for x is

$$t_{\text{hit}[\mathbf{P}]}(0,x) = \sum_{k \neq x} \mathrm{E}\left[Y^{(0)}_{0,k}\right] + \mathrm{E}\left[Y^{(1)}_{0,k}\right] + \mathrm{E}\left[Y^{(2)}_{0,k}\right] + ...$$

An similarly for y

$$t_{\text{hit}[\mathbf{P}]}(0,y) = \sum_{k \neq y} \mathrm{E}\left[Y^{(0)}_{0,k}\right] + \mathrm{E}\left[Y^{(1)}_{0,k}\right] + \mathrm{E}\left[Y^{(2)}_{0,k}\right] + ...$$

Prove

$$\sum_{k \neq x} \mathrm{E}\left[Y^{(0)}_{0,k}\right] + \mathrm{E}\left[Y^{(1)}_{0,k}\right] + \mathrm{E}\left[Y^{(2)}_{0,k}\right] + ... < \sum_{k \neq y} \mathrm{E}\left[Y^{(0)}_{0,k}\right] + \mathrm{E}\left[Y^{(1)}_{0,k}\right] + \mathrm{E}\left[Y^{(2)}_{0,k}\right] + ...$$

Noting as in Lemma 1 most terms cancel leaving

$$\mathrm{E}\left[Y^{(0)}_{0,y}\right] + \mathrm{E}\left[Y^{(1)}_{0,y}\right] + \mathrm{E}\left[Y^{(2)}_{0,y}\right] + ... < \mathrm{E}\left[Y^{(0)}_{0,x}\right] + \mathrm{E}\left[Y^{(1)}_{0,x}\right] + \mathrm{E}\left[Y^{(2)}_{0,x}\right] + ...$$

Using the definition of the expectation

$$\left[{}^{\alpha(x)}\mathbf{P}\right]^{0}_{0,y} + \left[{}^{\alpha(x)}\mathbf{P}\right]^{1}_{0,y} + \left[{}^{\alpha(x)}\mathbf{P}\right]^{2}_{0,y} + ... < \left[{}^{\alpha(y)}\mathbf{P}\right]^{0}_{0,x} + \left[{}^{\alpha(y)}\mathbf{P}\right]^{1}_{0,x} + \left[{}^{\alpha(y)}\mathbf{P}\right]^{2}_{0,x} + ...$$

Since for the states involved on both sides are transient states in their respective absorbing chains

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{0}_{0,y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{1}_{0,y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{2}_{0,y} + ... < \left[{}^{\alpha(y)}\mathbf{Q}\right]^{0}_{0,x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{1}_{0,x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{2}_{0,x} + ...$$

Which converge to the fundamental matrix $\mathbf{N}$ on both sides

$$\left[{}^{\alpha(x)}\mathbf{N}\right]_{0,y} < \left[{}^{\alpha(y)}\mathbf{N}\right]_{0,x}$$

Also note that for any term of those series (show for the $\alpha(x)$ side but symmetrical for $\alpha(y)$)

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i}_{0,y} = \sum_k \left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,k}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{k,y}$$

Since, y is only reachable from the state that represents its group g_y

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i}_{0,y} = \left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}$$

Thus the series would be

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{0}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{1}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{2}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} + ...$$

Allowing us to rewrite the converged series as (using the symmetrical argument on the $\alpha(y)$ side)

$$\left[{}^{\alpha(x)}\mathbf{N}\right]_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left[{}^{\alpha(y)}\mathbf{N}\right]_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

With out loss of generality let g_x and g_y be neighbors in the chain such that $\Pr\left[g_x \xrightarrow{1} g_y\right] > 0$. Since $s(x) > s(y)$ we know all paths from 0 to g_y must traverse g_x. Let n the minimum path length to the state g_x then

$$\forall_{k<n}\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^k_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}\right) = 0$$
$$\forall_{k<n}\left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^k_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}\right) = 0$$

This yields a new way to write the statement

$$\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+1}_{0,g_y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+2}_{0,g_y} + \ldots\right)\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}$$
$$<$$
$$\left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+2}_{0,g_x} + \ldots\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Breaking things up piecewise

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$
$$\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_y} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+1}_{0,g_y}\right)\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$
$$\ldots$$

Now for statement 0

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Since n is the length of the shortest path to g_x and g_y is only reachable through g_y

$$0 < \left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Proving statement 0. Since g_x and g_y are neighbors, for statement 1

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+1}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_x,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_x,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Since n is the minimum path length to g_x the process cannot have reached x (or y) in n steps thus

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_x} = \left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x}$$

And

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_x,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Using the construction of the chain

$$\frac{1}{2|L|}\frac{1}{2|g_y|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^n_{0,g_x} < \frac{1}{2|g_x|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^n_{0,g_x} + \frac{1}{2|g_x|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}$$

Since $|g_x| < |L|$ and $|g_y| < |L|$ statement 1 is proved.

For the rest of the statements we will use the inductive hypothesis that the following statement for $i-1$ is true

$$\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n}_{0,g_y}+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+1}_{0,g_y}+...+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_y}\right)\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}<\left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n}_{0,g_x}+\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}+...+\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Now prove the inductive case

$$\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n}_{0,g_y}+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+1}_{0,g_y}+...+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i}_{0,g_y}\right)\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}<\left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n}_{0,g_x}+\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,g_x}+...+\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i}_{0,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Rewriting left side (symmetrical for right)

$$\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n}_{0,g_y}+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+1}_{0,g_y}+...+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_y}\right)\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}$$

Using the hypothesis

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}<\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i}_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Break up the left side:

$$\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_x}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_x,g_y}+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_{(y+1)}}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_{(y+1)},g_y}+\sum_{l\in g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{l,g_y}\right)\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}$$

Note that by the construction of the chain

$$\forall_{l\in g_y}\left(\left[{}^{\alpha(x)}\mathbf{Q}\right]_{l,g_y}=1\right)$$

Simplifying (where l is any location in g_y) the inside becomes

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_x}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_x,g_y}+\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_{(y+1)}}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_{(y+1)},g_y}+|g_y|\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,l}$$

Now for the right side:

$$\left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_{(x-1)}}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_{(x-1)},g_x}+\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_y}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_y,g_x}+\sum_{m\in g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,m}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{m,g_x}\right)\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

Simplifying using similar observations the inside becomes

$$\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_{(x-1)}}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_{(x-1)},g_x}+\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_y}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_y,g_x}+|g_x|\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,m}$$

The last term on both sides can be rewritten (using the left side as an example):

$$|g_y|\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-2}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,l\in g_y}$$
$$\frac{|g_y|}{2|g_y|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-2}_{0,g_y}=\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-2}_{0,g_y}$$

Now putting things together into 3 inequalities

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_x}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_x,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}<\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_{(x-1)}}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_{(x-1)},g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-1}_{0,g_{(y+1)}}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_{(y+1)},g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}<\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-1}_{0,g_y}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_y,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$
$$\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{n+i-2}_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y}<\frac{1}{2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+i-2}_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x}$$

The first inequality is of the form (with $t\geq n$)

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g,(g+1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{(g+1),l\in(g+1)}<\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,(g-1)}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{(g-1),g}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g,l\in g}$$

The second inequality is of the form

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{(g+1),g}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g,l\in g}<\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,g}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g,(g-1)}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{(g-1),l\in(g-1)}$$

The last inequality is of the form

$$\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{(g+1),l\in(g+1)} < \frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{g,l\in g}$$

By construction transitions from a state representing a group g to

A location will be $\frac{1}{2}\frac{1}{|g|}$

Another group $g+1$ will be $\frac{1}{2}\frac{1}{|L|} = \frac{1}{2|L|}$

Another group $g-1$ will be $\frac{1}{2}\frac{|L|-1}{|L|} = \frac{|L|-1}{2|L|}$

Rewriting the first 2 inequalities

$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{(g+1),l\in(g+1)} < \frac{1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,(g-1)}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g,l\in g}$$
$$\frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g,l\in g} < \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,g}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{(g-1),l\in(g-1)}$$

Without loss of generality assume there are only 3 group states $(g-1)$, g, and $(g+1)$. To create a lower bound on the left side and upper bound on the right assume g and $(g+1)$ have 1 element and $(g-1)$ has the rest of the locations. This yields

$$\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g} < \frac{1}{2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,(g-1)}$$
$$\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)} < \frac{1}{2}\frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,g}$$

Canceling out terms yields the expressions which need to be proven.

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g} < \left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,(g-1)}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,g}$$

In the same way for the last inequality using the bounds established for the first two inequalities:

$$\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{(g+1),l\in(g+1)} < \frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{g,l\in g}$$
$$\frac{1}{2}\frac{1}{2|(g+1)|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)} < \frac{1}{2}\frac{1}{2|g|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}$$
$$\frac{1}{4|(g+1)|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)} < \frac{1}{4|g|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}$$

Which cancel since we are going to assume that g and $g+1$ have the same number of elements. Note the, the worst case was already covered by the second inequality above.

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)} < \left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,g}$$

By proving the middle inequality we can prove the others.

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{t}_{0,(g+1)} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{t}_{0,g}$$

Prove this general form to complete the lemma.

Prove the inequality on an induction on i. We are going to continue to assume that $(g-1)=0$ and that there are 3 states in the configuration described above.

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i}_{0,g} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i}_{0,(g-1)}$$

Let $i = 0$

$$0 < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{0}_{0,(g-1)} \quad \checkmark$$

Let $i = 1$

$$\begin{aligned}
\left[{}^{\alpha(x)}\mathbf{Q}\right]^{1}_{0,g} &< \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{1}_{0,(g-1)} \\
\sum_{k}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{0}_{0,k}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{k,g} &< \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{1}_{0,(g-1)} \\
\left[{}^{\alpha(x)}\mathbf{Q}\right]^{0}_{0,(g-1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{(g-1),g} &< \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,(g-1)} \\
\left[{}^{\alpha(x)}\mathbf{Q}\right]_{(g-1),g} &< \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,(g-1)}
\end{aligned}$$

Recall when the inequalities worst case bounds were chosen for the transitions out of each group. Inserting the appropriate bound gives:

$$\begin{aligned}
\frac{1}{2|L|} &< \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,(g-1)} \\
\frac{|L|-2}{2|L|} &< \left[{}^{\alpha(y)}\mathbf{Q}\right]^{n+1}_{0,(g-1)}
\end{aligned}$$

Breaking apart the right side gives

$$\begin{aligned}
&\sum_{k}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n}_{0,k}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{k,(g-1)} \\
&\left[{}^{\alpha(y)}\mathbf{Q}\right]^{n}_{0,(g-1)}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{(g-1),(g-1)}
\end{aligned}$$

The sum represents the fact that $g-1=0$ and has the chain takes the self loop to get to $(g-1)$. However, it also means $n=0$ and the second part of the sum is zero. So if $g-1=0$

$$\begin{aligned}
\frac{|L|-2}{2|L|} &< \mathbf{I}_{0,0}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,0} + 0 \\
\frac{|L|-2}{2|L|} &< \left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,0} \\
\frac{|L|-2}{2|L|} &< \frac{|L|-1}{2|L|} \\
|L|-2 &< |L|-1 \\
|L| &< |L|+1 \quad \checkmark
\end{aligned}$$

The inductive hypothesis is that the following statement is true

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)}$$

Now for the general case

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i}_{0,g} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i}_{0,(g-1)}$$

Breaking apart the left hand side

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i}_{0,g}$$
$$\sum_k \left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,k} \left[{}^{\alpha(x)}\mathbf{Q}\right]_{k,g}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} \left[{}^{\alpha(x)}\mathbf{Q}\right]_{g-1,g} + \left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} \left[{}^{\alpha(x)}\mathbf{Q}\right]_{g+1,g} + \sum_{l\in g} \left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,l} \left[{}^{\alpha(x)}\mathbf{Q}\right]_{l,g}$$
$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} + |g|\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,l}$$
$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} + |g|\sum_k \left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,k}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{k,l}$$
$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} + |g|\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,g}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g,l}$$
$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} + |g|\frac{1}{2|g|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,g}$$
$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} + \frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,g}$$

Breaking apart the right hand side

$$\frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i}_{0,(g-1)}$$
$$\frac{1}{|L|-2}\sum_k \left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,k}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{k,(g-1)}$$
$$\frac{1}{|L|-2}\left(\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{(g-1),(g-1)} + \left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g,(g-1)} + \sum_{l\in(g-1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,l}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{l,(g-1)}\right)$$
$$\frac{1}{|L|-2}\left(\frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g} + |(g-1)|\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,l}\right)$$
$$\frac{1}{|L|-2}\left(\frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g} + |(g-1)|\sum_k\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,k}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{k,l}\right)$$
$$\frac{1}{|L|-2}\left(\frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g} + |(g-1)|\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,(g-1)}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{(g-1),l}\right)$$
$$\frac{1}{|L|-2}\left(\frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g} + |(g-1)|\frac{1}{2|(g-1)|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,(g-1)}\right)$$
$$\frac{1}{|L|-2}\left(\frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g} + \frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,(g-1)}\right)$$

Now the formula as piecewise inequalities

$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} < \frac{|L|-1}{2|L|}\frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)}$$
$$\frac{|L|-1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} < \frac{|L|-1}{2|L|}\frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g}$$
$$\frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,g} < \frac{1}{2}\frac{1}{|L|-2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,(g-1)}$$

The common parts cancel

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g-1} < \frac{|L|-1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,(g-1)}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-1}_{0,g+1} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]^{i-1}_{0,g}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,g} < \frac{1}{|L|-2}\left[{}^{\alpha(x)}\mathbf{Q}\right]^{i-2}_{0,(g-1)}$$

For the first inequality in Lemma 1 we proved

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g-1}^{i-1} = \left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,(g-1)}^{i-1}$$

Thus,

$$|L| - 2 < |L| - 1 \quad \checkmark$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g+1}^{i-1} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,g}^{i-1}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g}^{i-2} < \frac{1}{|L|-2}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g-1)}^{i-2}$$

Applying the inductive hypothesis

$$1 < \frac{1}{|L|-2} \quad \checkmark$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g+1)}^{i-1} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,g}^{i-1}$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g}^{i-2} < \frac{1}{|L|-2}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g-1)}^{i-2} \quad \checkmark$$

Only the middle inequality remains, do a similar induction on $j > 1$. Noting $j = 1$

$$0 < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,g}^{1} \quad \checkmark$$

The general case

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g+1)}^{j} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,g}^{j}$$

Breaking down the left side yields (skipping the steps)

$$\frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g}^{j-1} + \frac{1}{2|L|}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g+1)}^{j-1} + \frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g+1)}^{j-2}$$

Breaking down the right side yields (skipping the steps)

$$\frac{1}{|L|-2}\left(\frac{1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,(g-1)}^{j-1} + \frac{|L|-1}{2|L|}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,(g+1)}^{j-1} + \frac{1}{2}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g}^{j-2}\right)$$

Piecewise, with common parts canceled

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g}^{j-1} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,(g-1)}^{j-1} \quad \checkmark$$
$$1 < \frac{|L|-1}{|L|-2} \quad \checkmark$$
$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g+1)}^{j-2} < \frac{1}{|L|-2}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g}^{j-2} \quad \checkmark$$

The two inductive hypothesises prove the first and last cases. This completes the induction to prove

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,(g+1)}^{t} < \frac{1}{|L|-2}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,g}^{t} \quad \checkmark$$

Which completes the lemma by proving

$$\left[{}^{\alpha(x)}\mathbf{Q}\right]_{0,g_y}^{t}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left[{}^{\alpha(y)}\mathbf{Q}\right]_{0,g_x}^{t}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x} \quad \checkmark$$
$$\left[{}^{\alpha(x)}\mathbf{N}\right]_{0,g_y}\left[{}^{\alpha(x)}\mathbf{Q}\right]_{g_y,y} < \left[{}^{\alpha(y)}\mathbf{N}\right]_{0,g_x}\left[{}^{\alpha(y)}\mathbf{Q}\right]_{g_x,x} \quad \checkmark$$

Therefore,

$$s(x) > s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) < t_{\text{hit}[\mathbf{P}]}(0,y)$$

□

Lemma 3. *For the chain in Definition 4 if location $x \in L$ has a smaller hitting time than location $y \in L$, $t_{hit[\mathbf{P}]}(0,x) < t_{hit[\mathbf{P}]}(0,y)$, then:*

$$s(x) > s(y)$$

Proof. Proof symmetrical to the proof for Lemma 2 when the statement is in contrapostive form:

$$s(x) \not> s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \not< t_{\text{hit}[\mathbf{P}]}(0,y)$$
$$s(x) \leq s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \geq t_{\text{hit}[\mathbf{P}]}(0,y)$$

By noting that via Lemma 1

$$s(x) = s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) t_{\text{hit}[\mathbf{P}]}(0,y) \quad \checkmark$$

Leaving only the symmetric proof to Lemma 2

$$s(x) < s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) > t_{\text{hit}[\mathbf{P}]}(0,y)$$
$$s(x) - s(y) < 0 \implies t_{\text{hit}[\mathbf{P}]}(0,x) - t_{\text{hit}[\mathbf{P}]}(0,y) > 0$$
$$-s(y) < -s(x) \implies -t_{\text{hit}[\mathbf{P}]}(0,y) > -t_{\text{hit}[\mathbf{P}]}(0,x)$$
$$s(y) > s(x) \implies t_{\text{hit}[\mathbf{P}]}(0,y) < t_{\text{hit}[\mathbf{P}]}(0,x) \quad \checkmark$$

Therefore

$$t_{\text{hit}[\mathbf{P}]}(0,x) < t_{\text{hit}[\mathbf{P}]}(0,y) \implies s(x) > s(y)$$

□

Lemma 4. *For the chain in Definition 4 if locations $x, y \in L$ have the same hitting time $t_{hit[\mathbf{P}]}(0,x) = t_{hit[\mathbf{P}]}(0,y)$ then:*

$$s(x) = s(y)$$

Proof. We will prove the contrapositive

$$s(x) \neq s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \neq t_{\text{hit}[\mathbf{P}]}(0,y)$$

Rewriting as two statements to get rid of the $\neq$ sign on the left

$$s(x) > s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \neq t_{\text{hit}[\mathbf{P}]}(0,y)$$
$$s(x) < s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \neq t_{\text{hit}[\mathbf{P}]}(0,y)$$

Lemma 2 proved

$$s(x) > s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) < t_{\text{hit}[\mathbf{P}]}(0,y)$$

Which proves

$$s(x) > s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \neq t_{\text{hit}[\mathbf{P}]}(0,y) \quad \checkmark$$

Lemma 3 proved

$$s(x) < s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) > t_{\text{hit}[\mathbf{P}]}(0,y)$$

Which proves

$$s(x) < s(y) \implies t_{\text{hit}[\mathbf{P}]}(0,x) \neq t_{\text{hit}[\mathbf{P}]}(0,y) \quad \checkmark$$

Therefore,

$$t_{\text{hit}[\mathbf{P}]}(0,x) \neq t_{\text{hit}[\mathbf{P}]}(0,y) \implies s(x) \neq s(y)$$

□

REFERENCES

[1] S. Ali, J. H. Andrews, T. Dhandapani, and W. Wang, "Evaluating the Accuracy of Fault Localization Techniques," *2009 IEEE/ACM International Conference on Automated Software Engineering*, pp. 76–87, 2009.

[2] C. M. Grinstead and J. L. Snell, *Introduction to Probability*, 2nd ed. Providence, RI: American Mathematical Society, 1997.