The Number of 2-Protected Nodes in Tries and PATRICIA Tries

Guan-Ru Yu

Department of Applied Mathematics
National Chiao Tung University

August 2nd, 2014

Three main classes of digital trees

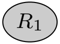
- Digital search trees (DSTs)
- Tries
- PATRICIA tries

Digital search trees (DSTs)

$$R_1 = 000001\cdots$$
$$R_2 = 000110\cdots$$
$$R_3 = 110111\cdots$$
$$R_4 = 011011\cdots$$
$$R_5 = 100001\cdots$$
$$R_6 = 111110\cdots$$

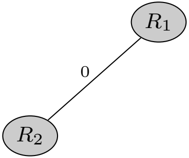
Digital search trees (DSTs)

$$R_1 = 000001\cdots$$
$$R_2 = 000110\cdots$$
$$R_3 = 110111\cdots$$
$$R_4 = 011011\cdots$$
$$R_5 = 100001\cdots$$
$$R_6 = 111110\cdots$$



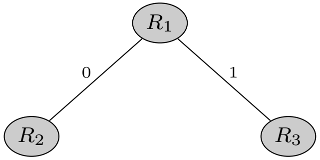
Digital search trees (DSTs)

$$R_1 = 000001\cdots$$
$$R_2 = 000110\cdots$$
$$R_3 = 110111\cdots$$
$$R_4 = 011011\cdots$$
$$R_5 = 100001\cdots$$
$$R_6 = 111110\cdots$$



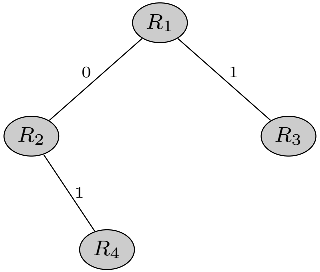
Digital search trees (DSTs)

$$R_1 = 000001\cdots$$
$$R_2 = 000110\cdots$$
$$R_3 = 110111\cdots$$
$$R_4 = 011011\cdots$$
$$R_5 = 100001\cdots$$
$$R_6 = 111110\cdots$$



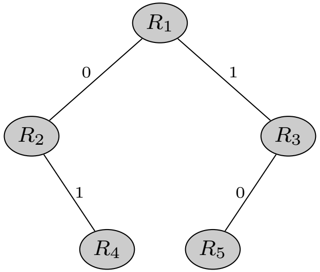
Digital search trees (DSTs)

$$R_1 = 000001\cdots$$
$$R_2 = 000110\cdots$$
$$R_3 = 110111\cdots$$
$$R_4 = 011011\cdots$$
$$R_5 = 100001\cdots$$
$$R_6 = 111110\cdots$$



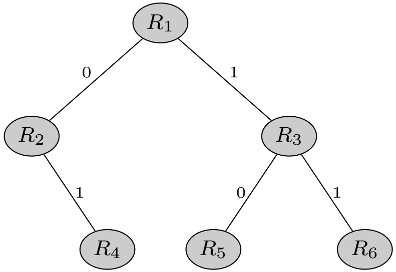
Digital search trees (DSTs)

$R_1 = 000001\cdots$
$R_2 = 000110\cdots$
$R_3 = 110111\cdots$
$R_4 = 011011\cdots$
$R_5 = 100001\cdots$
$R_6 = 111110\cdots$



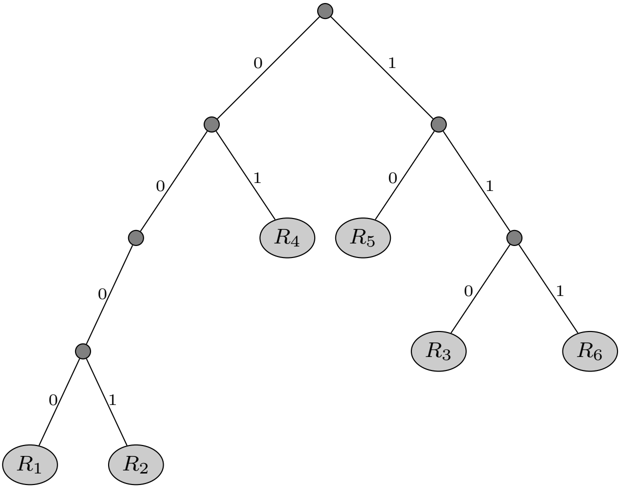
Digital search trees (DSTs)

$$R_1 = 000001\cdots$$
$$R_2 = 000110\cdots$$
$$R_3 = 110111\cdots$$
$$R_4 = 011011\cdots$$
$$R_5 = 100001\cdots$$
$$R_6 = 111110\cdots$$



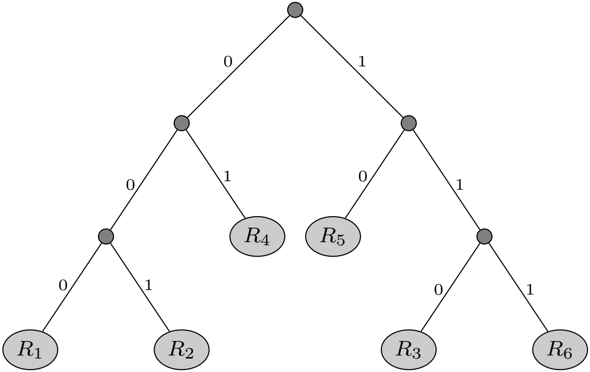
Tries

$R_1 = 000001\cdots$
$R_2 = 000110\cdots$
$R_3 = 110111\cdots$
$R_4 = 011011\cdots$
$R_5 = 100001\cdots$
$R_6 = 111110\cdots$



PATRICIA tries

$R_1 = 000001\cdots$
$R_2 = 000110\cdots$
$R_3 = 110111\cdots$
$R_4 = 011011\cdots$
$R_5 = 100001\cdots$
$R_6 = 111110\cdots$

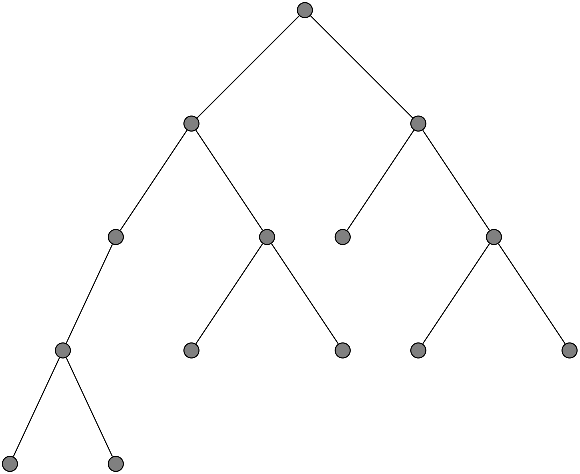


Random model

- (In probability theory) A sequence of RVs is independent and identically distributed (iid): same probability distribution and all are mutually independent.
- X_1, X_2, $\ldots$ is a random string: X_1, X_2, $\ldots$ is an iid sequence of RVs with $P(X_n = 0) = p$ and $P(X_n = 1) = q := 1 - p$.
- A digital tree is a random digital tree of size n: constructed from n infinite $\{0, 1\}$-random strings tries.

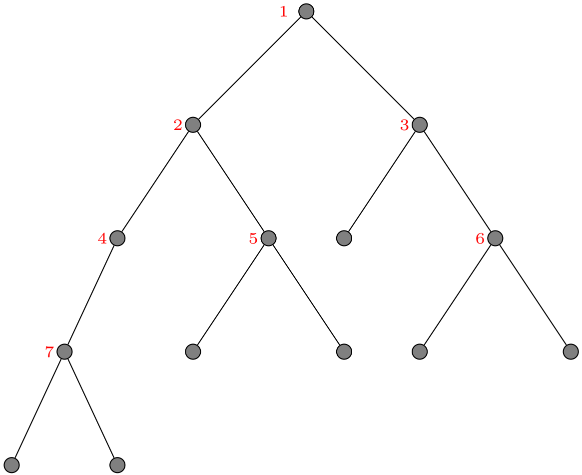
Size

Size of digital trees: number of internal nodes.



Size

Size of digital trees: number of internal nodes.



Size of tries

- Mean: Knuth (1973).

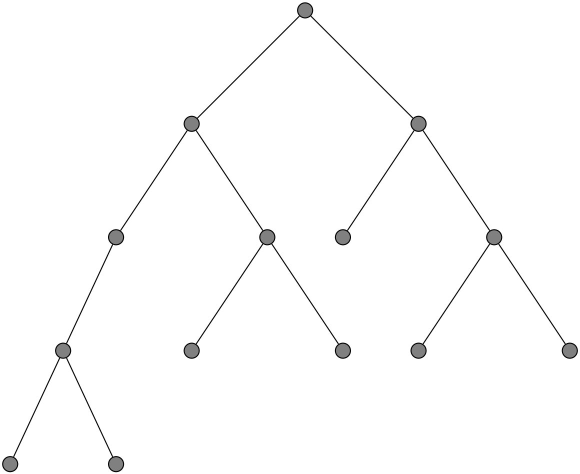
Size of tries

- Mean: Knuth (1973).
- Variance: Two groups independently.
- 1. P. Kirschenhofer and H. Prodinger: symmetric case ($p = q = 1/2$) with explicit expressions for involved constants and periodic functions (1991).
- 2. P. Jacquet and M. Regnier: general case (symmetric and asymmetric case) but without explicit expressions for involved constants and periodic functions (1988)

  partial results on explicit expressions of involved constants and periodic functions (1989).

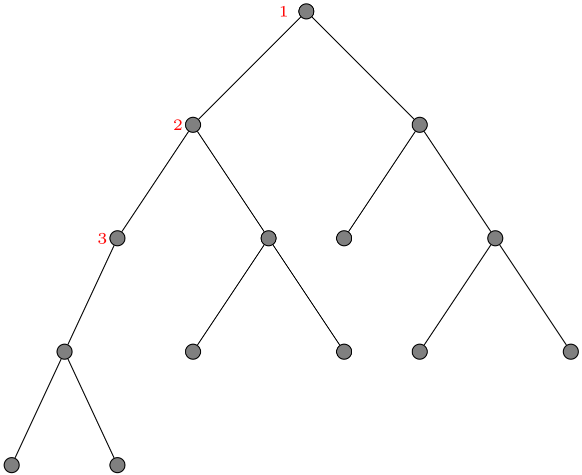
2-protected nodes

2-protected nodes: nodes that have distance at least 2 from leafs.



2-protected nodes

2-protected nodes: nodes that have distance at least 2 from leafs.



2-protected nodes in tries

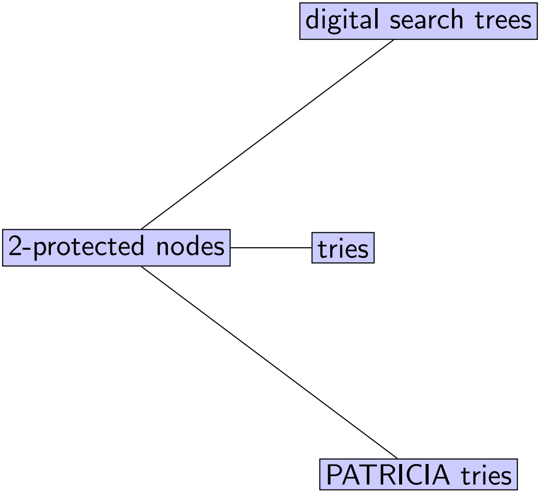
- Mean: J. Gaither, Y. Homma, M. Sellke and M. D. Ward (2012).

2-protected nodes in tries

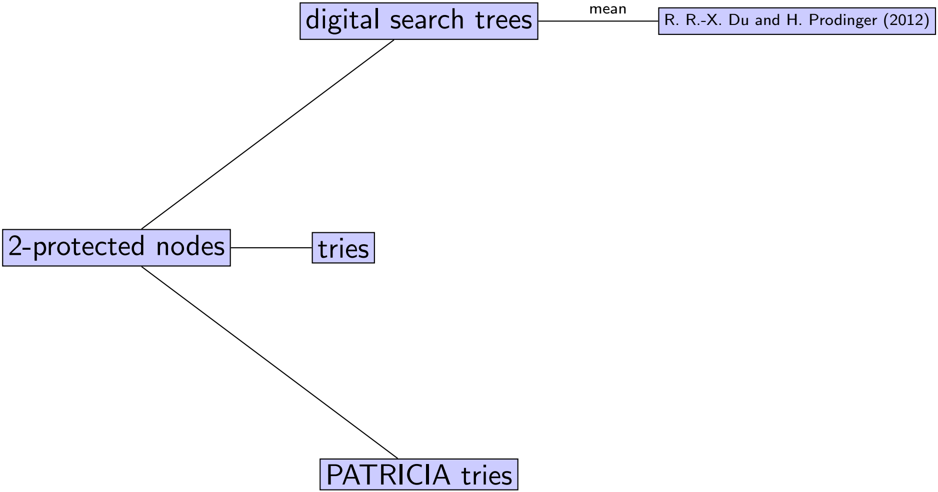
- Mean: J. Gaither, Y. Homma, M. Sellke and M. D. Ward (2012).
- Variance: J. Gaither and M. D. Ward (2013).
- Some mistakes:
  1. wrong error term.
  2. forgot to pull out the average value.

Number of 2-protected nodes in digital tree

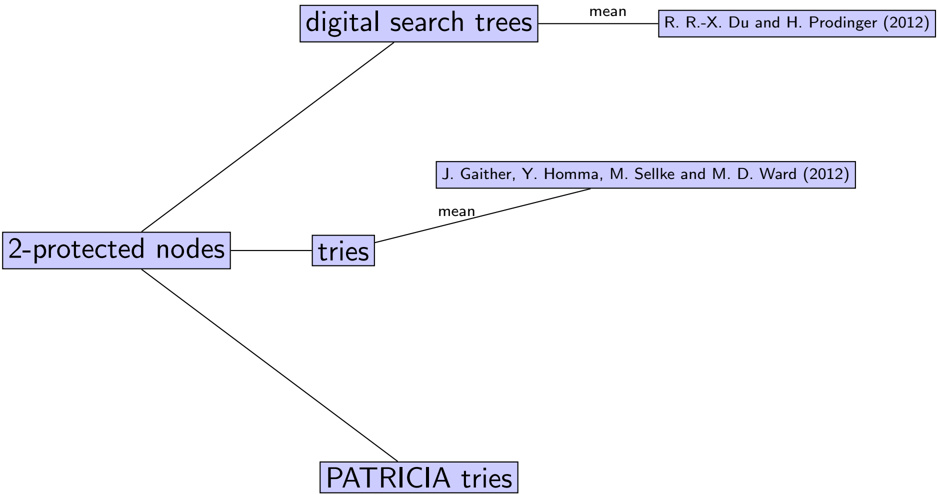
Number of 2-protected nodes in digital tree



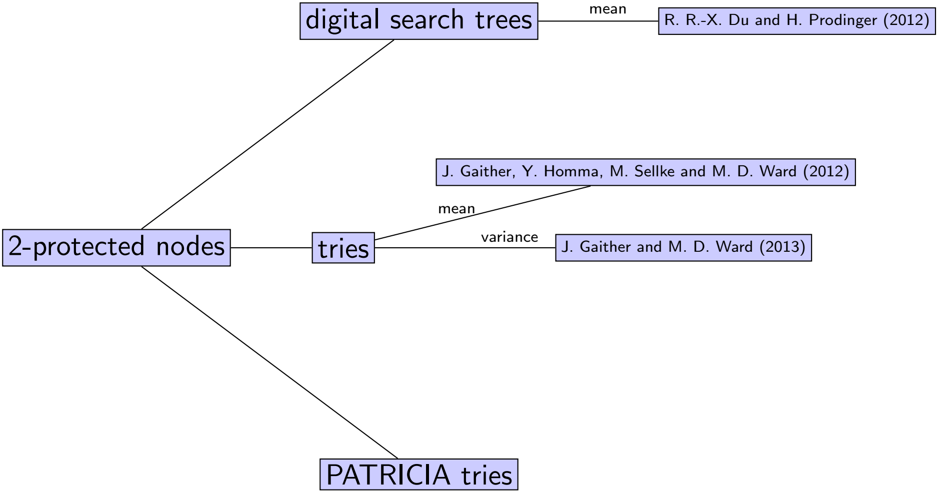
Number of 2-protected nodes in digital tree



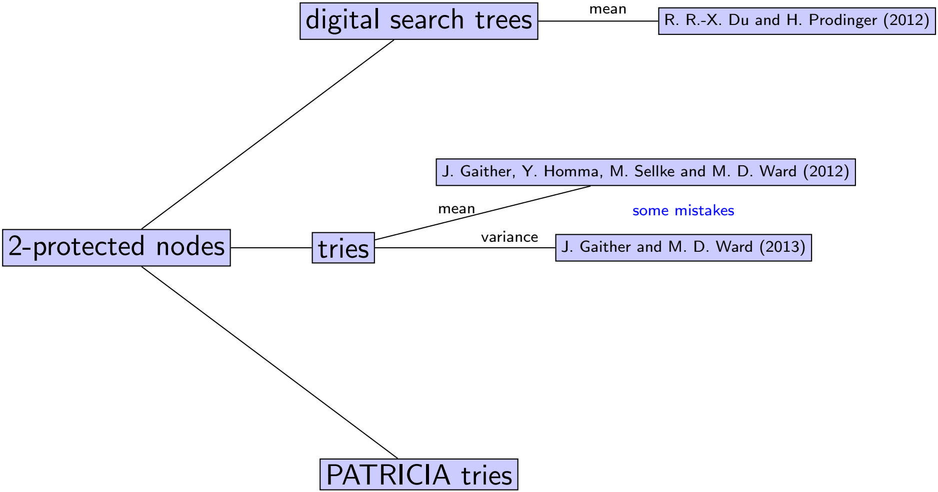
Number of 2-protected nodes in digital tree



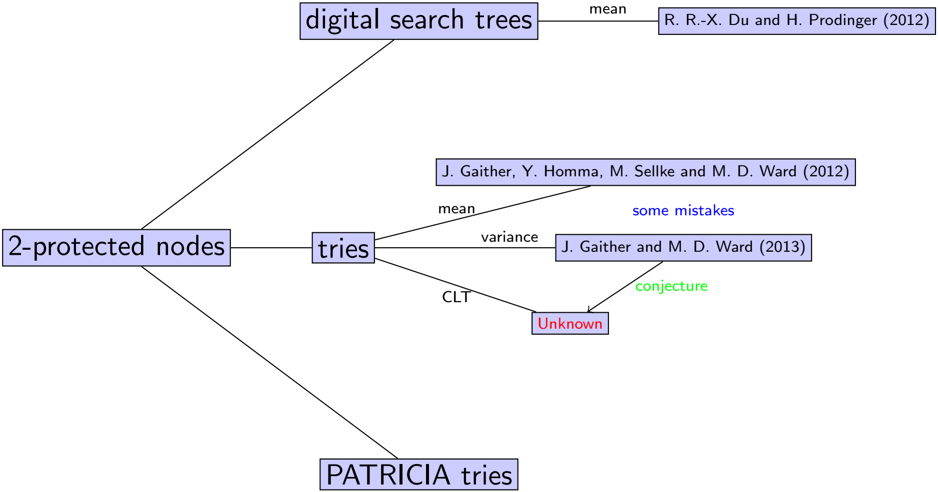
Number of 2-protected nodes in digital tree



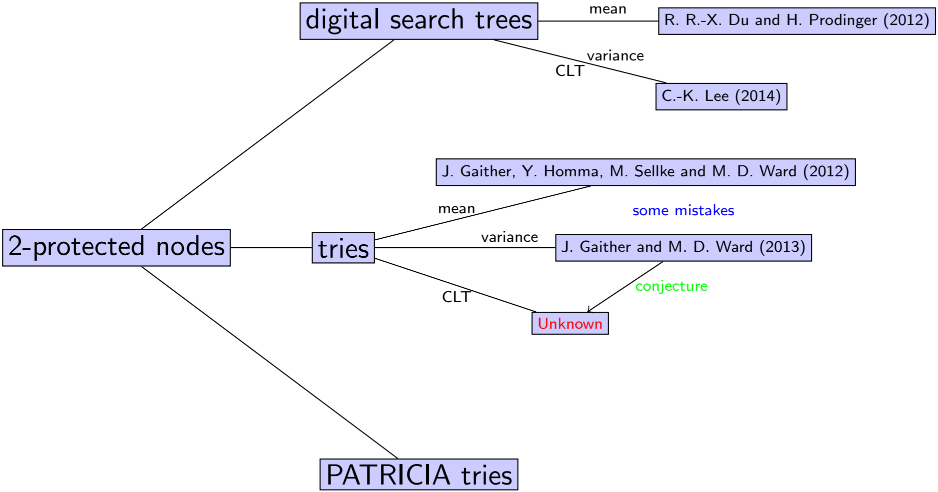
Number of 2-protected nodes in digital tree



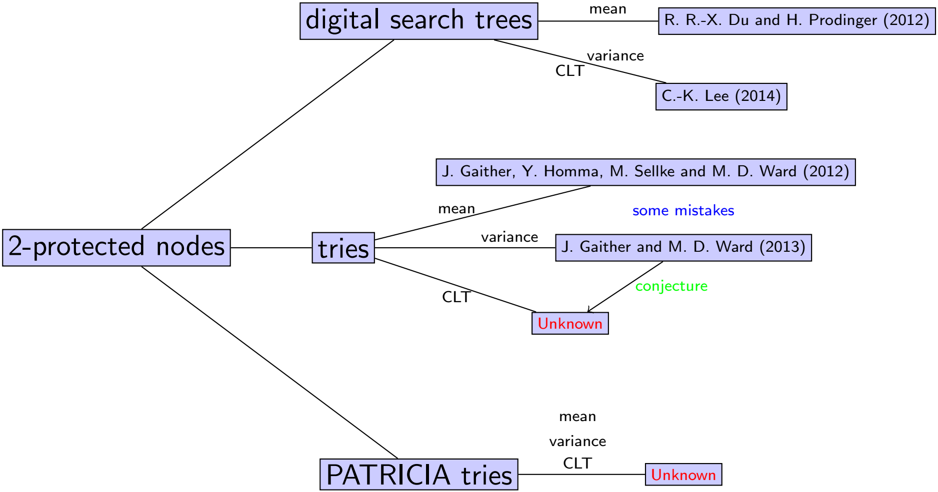
Number of 2-protected nodes in digital tree



Number of 2-protected nodes in digital tree



Number of 2-protected nodes in digital tree



Main goal of thesis

- Re-derive (and correcte) the results of J. Gaither, Y. Homma, M. Sellke and M. D. Ward by using the method of M. Fuchs, H.-K. Hwang, and V. Zacharovas.

Main goal of thesis

- Re-derive (and correcte) the results of J. Gaither, Y. Homma, M. Sellke and M. D. Ward by using the method of M. Fuchs, H.-K. Hwang, and V. Zacharovas.
- Prove the conjectured central limit theorem.

Main goal of thesis

- Re-derive (and correcte) the results of J. Gaither, Y. Homma, M. Sellke and M. D. Ward by using the method of M. Fuchs, H.-K. Hwang, and V. Zacharovas.
- Prove the conjectured central limit theorem.
- Derive a bivariate central limit theorem of the number of internal nodes and the number of 2-protected nodes in random tries. This result contains the central limit theorem for PATRICIA tries.

Main goal of thesis

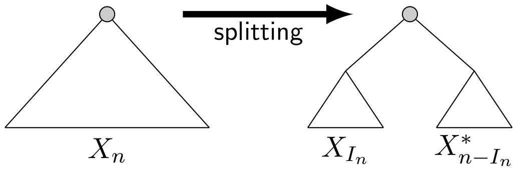
- Re-derive (and correcte) the results of J. Gaither, Y. Homma, M. Sellke and M. D. Ward by using the method of M. Fuchs, H.-K. Hwang, and V. Zacharovas.
- Prove the conjectured central limit theorem.
- Derive a bivariate central limit theorem of the number of internal nodes and the number of 2-protected nodes in random tries. This result contains the central limit theorem for PATRICIA tries.
- Derive asymptotic expansions for the number of 2-protected nodes in PATRICIA tries.

Additive shape parameters

- A general framework for obtaining asymptotic expansions of mean and variance of additive shape parameter in random tries with explicit expressions for periodic functions.

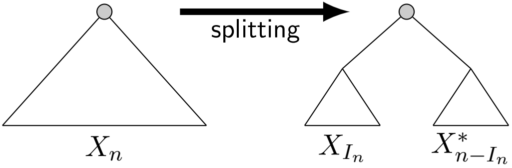
Additive shape parameters

- A general framework for obtaining asymptotic expansions of mean and variance of additive shape parameter in random tries with explicit expressions for periodic functions.



Additive shape parameters

- A general framework for obtaining asymptotic expansions of mean and variance of additive shape parameter in random tries with explicit expressions for periodic functions.



- By splitting the trie,

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n.$$

Main steps of the analysis

- The first step is to take moments on both sides of the distributional recurrence for X_n. This gives

$$a_n = \sum_{0 \le k \le n} \pi_{n,k}(a_k + a_{n-k}) + b_n$$

for all moments, where b_n is a function of moments of lower order.

Main steps of the analysis

- The first step is to take moments on both sides of the distributional recurrence for X_n. This gives

$$a_n = \sum_{0\le k\le n} \pi_{n,k}(a_k + a_{n-k}) + b_n$$

for all moments, where b_n is a function of moments of lower order.

- The second step is poissonization. Let

$$\tilde{f}(z) := e^{-z}\sum_n a_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z}\sum_n b_n \frac{z^n}{n!}.$$

Main steps of the analysis

- The first step is to take moments on both sides of the distributional recurrence for X_n. This gives

$$a_n = \sum_{0\leq k\leq n} \pi_{n,k}(a_k + a_{n-k}) + b_n$$

for all moments, where b_n is a function of moments of lower order.

- The second step is poissonization. Let

$$\tilde{f}(z) := e^{-z}\sum_n a_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z}\sum_n b_n \frac{z^n}{n!}.$$

- We have

$$\tilde{f}(z) = \tilde{f}(pz) + \tilde{f}(qz) + \tilde{g}(z).$$

Main steps of the analysis

- The third step is doing Mellin transform on both sides. Therefore

$$\mathscr{M}[\tilde{f}(z); s] := \int_0^\infty \tilde{f}(z) z^{s-1} \mathrm{d}z.$$

Main steps of the analysis

- The third step is doing Mellin transform on both sides. Therefore

$$\mathscr{M}[\tilde{f}(z); s] := \int_0^\infty \tilde{f}(z) z^{s-1} \mathrm{d}z.$$

- We have

$$\mathscr{M}[\tilde{f}(z); s] = \frac{\mathscr{M}[\tilde{g}(z); s]}{1 - p^{-s} - q^{-s}}$$

Main steps of the analysis

- The third step is doing Mellin transform on both sides. Therefore

$$\mathscr{M}[\tilde{f}(z); s] := \int_0^\infty \tilde{f}(z) z^{s-1} \mathrm{d}z.$$

- We have

$$\mathscr{M}[\tilde{f}(z); s] = \frac{\mathscr{M}[\tilde{g}(z); s]}{1 - p^{-s} - q^{-s}}$$

- The fourth step is using inverse Mellin transform

$$\tilde{f}(z) = \frac{1}{2\pi i} \int_{\uparrow} \frac{\mathscr{M}[\tilde{g}(z); s] z^{-s}}{1 - p^{-s} - q^{-s}} ds$$

and applying the converse mapping theorem.

Main steps of the analysis

- The last step is depoissonization which is based on the Poisson heuristic

$$\tilde{f}(n) = \mathbb{E}(a_N) \sim a_n.$$

Main steps of the analysis

- The last step is depoissonization which is based on the Poisson heuristic

$$\tilde{f}(n) = \mathbb{E}(a_N) \sim a_n.$$

- To make this step precise, we use Cauchy's integral formula

$$a_n = \frac{n!}{2\pi i} \oint_{|z|=r} z^{-n-1} e^z \tilde{f}(z) dz$$

and the saddle-point method.

Main steps of the analysis

- The last step is depoissonization which is based on the Poisson heuristic

$$\tilde{f}(n) = \mathbb{E}(a_N) \sim a_n.$$

- To make this step precise, we use Cauchy's integral formula

$$a_n = \frac{n!}{2\pi i} \oint_{|z|=r} z^{-n-1} e^z \tilde{f}(z) dz$$

and the saddle-point method.

- All the above steps can be merged via

$$a_n = \frac{n!}{2\pi i} \int_{\uparrow} \frac{\mathscr{M}[\tilde{g}(z); s]}{(1 - p^{-s} - q^{-s})\Gamma(n+1-s)} ds.$$

Poisson-Mellin-Newton cycle

a_n

Recurrence relation

$$a_n = \sum_{0 \le k \le n} \pi_{n,k}(a_k + a_{n-k}) + b_n$$

Poisson generating function

$$\tilde{f}(z) = \tilde{f}(pz) + \tilde{f}(qz) + \tilde{g}(z)$$

Mellin transform

$$\mathscr{M}[\tilde{f}(z); s] = \frac{\mathscr{M}[\tilde{g}(z); s]}{1 - p^{-s} - q^{-s}}$$

Inverse Mellin transform

$$\tilde{f}(z) = \frac{1}{2\pi i} \int_{\uparrow} \frac{\mathscr{M}[\tilde{g}(z); s] z^{-s}}{1 - p^{-s} - q^{-s}} ds$$

Analytic de-Poissonization

$$a_n = \frac{n!}{2\pi i} \oint_{|z|=r} z^{-n-1} e^z \tilde{f}(z) dz$$

Variance

- The definition of the variance is

$$\mathbb{V}(X_n) = \mathbb{E}(X_n^2) - (\mathbb{E}(X_n))^2.$$

Variance

- The definition of the variance is

$$\mathbb{V}(X_n) = \mathbb{E}(X_n^2) - (\mathbb{E}(X_n))^2.$$

- P. Jacquet and M. Regnier used the following

$$\tilde{W}(z) := \tilde{f}_2(z) - \tilde{f}_1(z)^2$$

Variance

- The definition of the variance is

$$\mathbb{V}(X_n) = \mathbb{E}(X_n^2) - (\mathbb{E}(X_n))^2.$$

- P. Jacquet and M. Regnier used the following

$$\tilde{W}(z) := \tilde{f}_2(z) - \tilde{f}_1(z)^2$$

- M. Fuchs, H.-K. Hwang and V. Zacharovas pointed out that a better choice is

$$\tilde{V}(z) := \tilde{f}_2(z) - \tilde{f}_1(z)^2 - z\tilde{f}_1'(z)^2.$$

We will make use of this in our analysis of 2-protected nodes in tries and PATRICIA tries.

2-protected nodes in tries

- Recall: X_n can be described by

$$X_n \overset{d}{=} X_{I_n} + X^*_{n-I_n} + T_n.$$

2-protected nodes in tries

- Recall: X_n can be described by

$$X_n \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n.$$

- By splitting the trie,

$$X_n^{(T)} \stackrel{d}{=} \begin{cases} X_{n-1}^{(T)}, & \text{when } I_n = 1 \text{ or } I_n = n-1; \\ X_{I_n}^{(T)} + X_{n-I_n}^{(T)*} + 1, & \text{otherwise,} \end{cases}$$

with initial conditions $X_0^{(T)} = X_1^{(T)} = 0$. So,

$$T_n = \begin{cases} 0, & \text{when } I_n = 1 \text{ or } I_n = n-1; \\ 1, & \text{otherwise.} \end{cases}$$

Mean of 2-protected nodes in tries

- With $\chi_k = \frac{2rk\pi i}{\log p}$ and $\frac{\log p}{\log q} = \frac{r}{l}$ and $(r, l) = 1$, let

$$\mathscr{F}[G](x) := \begin{cases} \dfrac{1}{h} \displaystyle\sum_{k \in \mathbb{Z} \setminus \{0\}} G(-1 + \chi_k) e^{2k\pi i x}, & \text{if } \frac{\log p}{\log q} \in \mathbb{Q}; \\ 0, & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q}, \end{cases}$$

Mean of 2-protected nodes in tries

- With $\chi_k = \frac{2rk\pi i}{\log p}$ and $\frac{\log p}{\log q} = \frac{r}{l}$ and $(r, l) = 1$, let

$$\mathscr{F}[G](x) := \begin{cases} \dfrac{1}{h} \displaystyle\sum_{k \in \mathbb{Z} \setminus \{0\}} G(-1 + \chi_k) e^{2k\pi i x}, & \text{if } \frac{\log p}{\log q} \in \mathbb{Q}; \\ 0, & \text{if } \frac{\log p}{\log q} \notin \mathbb{Q}, \end{cases}$$

Theorem

We have

$$\frac{\mathbb{E}(X_n^{(T)})}{n} = \frac{pq + 1 - h}{h} + \mathscr{F}[G_1^{(T)}](r \log_{1/p} n) + o(1),$$

where $G_1^{(T)}(-1 + \chi_k) = \Gamma(-1 + \chi_k)\chi_k(\chi_k pq - pq - 1)$.

Variance of 2-protected nodes in tries

Theorem

We have

$$\frac{\mathbb{V}(X_n^{(T)})}{n} = \frac{G_2^{(T)}(-1)}{h} + \mathscr{F}[G_2^{(T)}](r\log_{1/p} n) + o(1).$$

where $G_2^{(T)}(-1+\chi_k)$ *is*

$$\begin{aligned}
&\Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right) + \Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k} + 2pq^2(1+q)^{-2-\chi_k}\right)\\
&+\Gamma(\chi_k+1)\left(-3pq + 2^{-\chi_k}pq - 2^{-1-\chi_k}\right) + \Gamma(\chi_k)\left(1-2p(1+p)^{-\chi_k} - 2q(1+q)^{-\chi_k}\right)\\
&+\Gamma(\chi_k-1)\left(1-2^{1-\chi_k}\right) - \frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(T)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(T)}(-1+\chi_{k-j})\\
&+2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k)
\end{aligned}$$

Variance of 2-protected nodes in tries

Theorem (continued)

$$+2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Big(pq(p^\ell+q^\ell)(\ell-1)\Gamma(\chi_k+\ell+1)+\Big(1-\ell(p^\ell+q^\ell)(2pq+p^{-\ell+2}$$

$$+q^{-\ell+2})\Big)\Gamma(\chi_k+\ell)+(p^\ell+q^\ell)(1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell)\Gamma(\chi_k+\ell-1)\Big)$$

and $G_2^{(T)}(-1)$ *is*

$$\frac{2p^2q}{(1+p)^2}+\frac{2pq^2}{(1+q)^2}-2pq-\frac{p^2q^2}{4}+2p\log(1+p)+2q\log(1+q)+\frac{1}{2}+h-2\log 2$$

$$+2\sum_{\ell\geq 2}(-1)^\ell\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Bigg((p^\ell+q^\ell)\Bigg(\frac{1-\ell+p^{-\ell+2}\ell+q^{-\ell+2}\ell}{\ell(\ell-1)}+pq\ell-3pq-p^{-\ell+2}$$

Variance of 2-protected nodes in tries

Theorem (continued)

$$-q^{-\ell+2}\Big)+\frac{1}{\ell}\Big)+2\sum_{\ell\geq 1}\frac{(-1)^{\ell}}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell)-\frac{1}{h}(pq+1-h)^2$$

$$-\begin{cases}\dfrac{1}{h\log p}\displaystyle\sum_{j\geq 1}\frac{4rj\pi^2}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2+(pq+1)^2\right), & \textit{if } \frac{\log p}{\log q}\in\mathbb{Q};\\ 0, & \textit{if } \frac{\log p}{\log q}\notin\mathbb{Q}.\end{cases}$$

Here, we set

$$K_1(s) := -1+pqs(s+1)-p^{-s}s-q^{-s}s$$

Variance of 2-protected nodes in tries

Theorem (continued)

$$-q^{-\ell+2}\Big)+\frac{1}{\ell}\Big)+2\sum_{\ell\geq 1}\frac{(-1)^{\ell}}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell)-\frac{1}{h}(pq+1-h)^2$$

$$-\begin{cases}\dfrac{1}{h\log p}\displaystyle\sum_{j\geq 1}\frac{4rj\pi^2}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2+(pq+1)^2\right), & \textit{if } \frac{\log p}{\log q}\in\mathbb{Q};\\ 0, & \textit{if } \frac{\log p}{\log q}\notin\mathbb{Q}.\end{cases}$$

Here, we set

$$K_1(s):=-1+pqs(s+1)-p^{-s}s-q^{-s}s$$

- Our expression is different from J. Gaither and M. D. Ward.

Variance of 2-protected nodes in tries

Theorem (continued)

$$- q^{-\ell+2}\Big) + \frac{1}{\ell}\Big) + 2\sum_{\ell\geq 1}\frac{(-1)^{\ell}}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell) - \frac{1}{h}(pq+1-h)^2$$

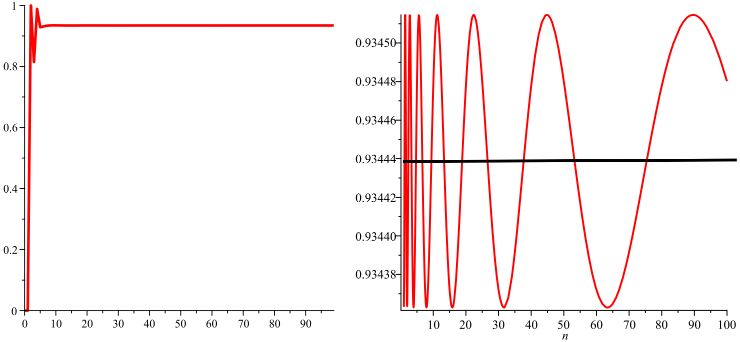
$$-\begin{cases} \dfrac{1}{h\log p}\displaystyle\sum_{j\geq 1}\frac{4rj\pi^2}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2 + (pq+1)^2\right), & \text{if } \frac{\log p}{\log q}\in\mathbb{Q}; \\ 0, & \text{if } \frac{\log p}{\log q}\notin\mathbb{Q}. \end{cases}$$

Here, we set

$$K_1(s) := -1 + pqs(s+1) - p^{-s}s - q^{-s}s$$

- Our expression is different from J. Gaither and M. D. Ward.
- We give an explanation why Abel summability gives the correct result in the expression above.

$\mathbb{V}(X_n^{(T)})/n$



Central limit theorem

Theorem (CLT)

We have,

$$\frac{X_n^{(T)} - \mathbb{E}(X_n^{(T)})}{\sqrt{\mathbb{V}(X_n^{(T)})}} \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ *denotes convergence in distribution and* $N(0,1)$ *is the standard normal distribution.*

2-protected nodes in PATRICIA tries

- By splitting the PATRICIA trie,

$$X_n^{(P)} \stackrel{d}{=} \begin{cases} X_n^{(P)}, & \text{when } I_n = 0 \text{ or } n; \\ X_{n-1}^{(P)}, & \text{when } I_n = 1 \text{ or } n-1; \\ X_{I_n}^{(P)} + X_{n-I_n}^{(P)*} + 1, & \text{otherwise,} \end{cases}$$

with initial conditions $X_0^{(P)} = X_1^{(P)} = 0$. So,

$$T_n = \begin{cases} 0, & \text{if } I_n \in \{0, 1, n-1, n\}; \\ 1, & \text{otherwise.} \end{cases}$$

2-protected nodes in PATRICIA tries

- By splitting the PATRICIA trie,

$$X_n^{(P)} \stackrel{d}{=} \begin{cases} X_n^{(P)}, & \text{when } I_n = 0 \text{ or } n; \\ X_{n-1}^{(P)}, & \text{when } I_n = 1 \text{ or } n-1; \\ X_{I_n}^{(P)} + X_{n-I_n}^{(P)*} + 1, & \text{otherwise,} \end{cases}$$

with initial conditions $X_0^{(P)} = X_1^{(P)} = 0$. So,

$$T_n = \begin{cases} 0, & \text{if } I_n \in \{0, 1, n-1, n\}; \\ 1, & \text{otherwise.} \end{cases}$$

- By the same analysis as for tries, we get the results of PATRICIA tries.

Mean of 2-protected nodes in PATRICIA tries

Mean of 2-protected nodes in PATRICIA tries

Theorem

We have

$$\frac{\mathbb{E}(X_n^{(P)})}{n} = \frac{pq}{h} + \mathscr{F}[G_1^{(P)}](r\log_{1/p} n) + o(1),$$

where $G_1^{(P)}(-1+\chi_k) = pq\Gamma(\chi_k + 1)$.

- Here, we have the following relationship between 2-protected nodes in tries and PATRICIA tries via the number of internal nodes:

$$X_n^{(P)} = X_n^{(T)} - N_n + n - 1.$$

Variance of 2-protected nodes in PATRICIA tries

Theorem

We have

$$\frac{\mathbb{V}(X_n^{(P)})}{n} = \frac{G_2^{(P)}(-1)}{h} + \mathscr{F}[G_2^{(P)}](r\log_{1/p} n) + o(1),$$

where $G_2^{(P)}(-1+\chi_k)$ *is*

$$\Gamma(3+\chi_k)\left(-2^{-3-\chi_k}p^2q^2\right) + \Gamma(\chi_k+2)\left(2p^2q(1+p)^{-2-\chi_k} + 2pq^2(1+q)^{-2-\chi_k} - 2^{-1-\chi_k}pq\right)$$
$$+\Gamma(\chi_k+1)\big(-3pq - 2^{-1-\chi_k}(1+2pq+p^{1-\chi_k}+q^{1-\chi_k}) + 2p(1+q)(1+p)^{-1-\chi_k}$$
$$+2q(1+p)(1+q)^{-1-\chi_k}\big) + (\Gamma(\chi_k)+\Gamma(\chi_k-1))\big(2(1+p)^{1-\chi_k} + 2(1+q)^{1-\chi_k} - 2^{1-\chi_k}$$
$$+(1-2^{1-\chi_k})(p^{1-\chi_k}+q^{1-\chi_k}) - 3\big) + 2\sum_{\ell\geq 2}\frac{(-1)^\ell}{\ell!}\frac{K_1(-\ell)}{1-p^\ell-q^\ell}\Big(pq(p^\ell+q^\ell)(\ell-1)$$
$$\Gamma(\chi_k+\ell+1) + \left((p^\ell+q^\ell)(\ell-2pq\ell-p^{-\ell+2}\ell-q^{-\ell+2}\ell-1)+1\right)\Gamma(\chi_k+\ell)$$

Variance of 2-protected nodes in PATRICIA tries

Theorem (continued)

$$+(1-p^\ell-q^\ell)\Gamma(\chi_k+\ell-1)\Big)-\frac{1}{h}\sum_{j\in\mathbb{Z}}(\chi_j-1)G_1^{(P)}(\chi_j-1)(-1+\chi_{k-j})G_1^{(P)}(-1+\chi_{k-j})$$

$$+2\sum_{\ell\geq 1}\frac{(-1)^\ell}{\ell!}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell+\chi_k-1)K_1(-1-\ell)\Gamma(\ell+\chi_k),$$

and $G_2^{(P)}(-1)$ *is*

$$-\frac{p^2q^2}{4}-\frac{9pq}{2}-1+\frac{2p(1+q)}{1+p}+\frac{2q(1+p)}{1+q}+\frac{2p^2q}{(1+p)^2}+\frac{2pq^2}{(1+q)^2}$$
$$+2\sum_{\ell\geq 2}\frac{(-1)^\ell K_1(-\ell)}{1-p^\ell-q^\ell}\left((p^\ell+q^\ell)\left(1-p^{-\ell+2}-q^{-\ell+2}-\frac{1}{\ell-1}+pq\ell-3pq\right)\right.$$

Variance of 2-protected nodes in PATRICIA tries

Theorem (continued)

$$+\frac{1}{\ell-1}\Bigg)+2\sum_{\ell\geq 1}\frac{(-1)^{\ell}}{\ell}\frac{p^{1+\ell}+q^{1+\ell}}{1-p^{1+\ell}-q^{1+\ell}}K_1(\ell-1)K_1(-1-\ell)-\frac{p^2q^2}{h}$$

$$-\begin{cases}\dfrac{1}{h\log p}\displaystyle\sum_{j\geq 1}\frac{4rj\pi^2}{\sinh\left(\frac{2rj\pi^2}{\log p}\right)}\left(p^2q^2\left(\frac{2rj\pi}{\log p}\right)^2+p^2q^2\right), & \textit{if } \frac{\log p}{\log q}\in\mathbb{Q};\\ 0, & \textit{if } \frac{\log p}{\log q}\notin\mathbb{Q}.\end{cases}$$

Here, we set $K_1(s) := (pqs+1-p^{-s}-q^{-s})(s+1)$.

Bivariate central limit theorem

Theorem

We have

$$\frac{\mathrm{Cov}(N_n, X_n^{(T)})}{n} = \frac{H_2(-1)}{h} + \mathscr{F}[H_2](r\log_{1/p} n) + o(1),$$

where

$$H_2(x) = \frac{G_2^{(T)}(x) + G_2^{(N)}(x) - G_2^{(P)}(x)}{2}.$$

Bivariate central limit theorem

Theorem

We have

$$\frac{\mathrm{Cov}(N_n, X_n^{(T)})}{n} = \frac{H_2(-1)}{h} + \mathscr{F}[H_2](r \log_{1/p} n) + o(1),$$

where

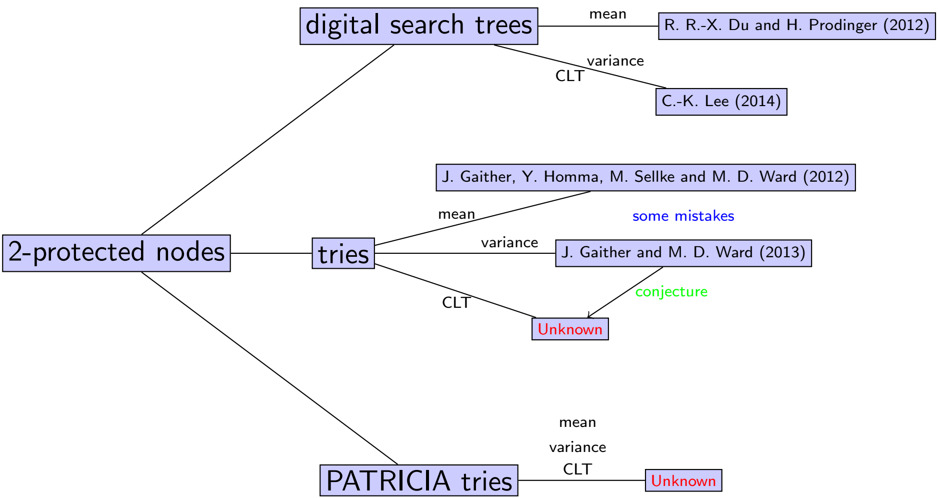
$$H_2(x) = \frac{G_2^{(T)}(x) + G_2^{(N)}(x) - G_2^{(P)}(x)}{2}.$$

Theorem

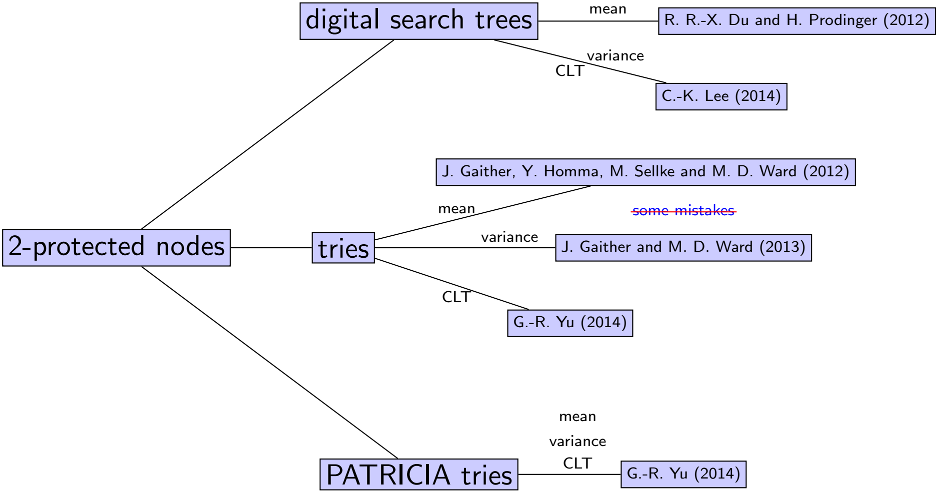
We have,

$$\Sigma_n^{-1/2} \begin{pmatrix} N_n - \mathbb{E}(N_n) \\ X_n^{(T)} - \mathbb{E}(X_n^{(T)}) \end{pmatrix} \xrightarrow{d} N(0, I_2),$$

Number of 2-protected nodes in digital tree



Number of 2-protected nodes in digital tree



Thanks for your attention!