Appendices

A DETAILED ANALYSES

A.1 Derivation of Expectation of Ψ for $\sigma' = Id \circ \sigma$

We use Indicator random variables $X_{i,j}$ to determine whether the join of $i \in R, j \in S$ is made.

$$X_{i,j} = \begin{cases} 0 & i, j \text{ not joined in } \Psi \\ 1 & i, j \text{ joined in } \Psi \end{cases}$$

The tuples are probed as long as they are stored.

$$E[X_{i,j}] = I[i.k = j.k] \cdot p(p_R q_s) = \frac{\varepsilon_r \varepsilon_s}{p} I[i.k = j.k]$$

The complete derivation is given by:

$$E[|\Psi|] = \sum E[X_{i,j}] = \sum_{i.k=j.k} E[X_{i,j}] = \frac{\varepsilon_R \varepsilon_S}{p} \gamma_{1,1} = \frac{\varepsilon_R \varepsilon_S}{p} J$$

A.2 Proof of Variance of $\hat{J}$ for $\sigma' = Id \circ \sigma$

We still use the notations in Appendix A.1. Obviously,

$$\mathrm{Var}[\hat{J}] = \frac{p^2}{(\varepsilon_R \varepsilon_S)^2} \mathrm{Var}[|\Psi|]$$

And we can obtain the result by calculating $\mathrm{Var}[|\Psi|]$.

$$\begin{aligned} \mathrm{Var}[|\Psi|] &= \sum \mathrm{Cov}(X_{r,s}, X_{r',s'}) \\ &= \sum E[X_{r,s}, X_{r',s'}] - E[X_{r,s}]E[X_{r',s'}] \end{aligned}$$

There is a case where there are two tuples a, b with the same key in R, S respectively. a and b are joined if and only if a is sampled by σ_R and b is sampled by σ_S.

$$\begin{aligned} & E[X_{r,s}, X_{r',s'}](r \neq r', s \neq s', r.k \neq r'.k) \\ & = E[X_{r,s}]E[X_{r',s'}] \\ & E[X_{r,s}, X_{r',s'}](r \neq r', s \neq s', r.k = r'.k) \\ & = pq_R^2 q_S^2 I[r.k = s.k = r'.k = s'.k] \\ & E[X_{r,s}, X_{r,s'}](s \neq s') \\ & = pq_R q_S^2 I[r.k = s.k = s'.k] \\ & E[X_{r,s}, X_{r',s}](r \neq r') \\ & = pq_R^2 q_S I[r.k = s.k = r'.k] \\ & E[X_{r,s}, X_{r,s}] = E[X_{r,s}] \end{aligned}$$

So the final variance is given by:

$$\begin{aligned} & \mathrm{Var}[\hat{J}] \\ & = \frac{p^2}{(\varepsilon_R \varepsilon_S)^2} \mathrm{Var}[|\Psi|] \\ & = \frac{p^2}{(\varepsilon_R \varepsilon_S)^2} \Big((\gamma_{2,2} - \gamma_{2,1} - \gamma_{1,2} + \gamma_{1,1})(pq_R^2 q_S^2 - p^2 p_R^2 q_s^2) \\ & \quad + \gamma_{1,1}(p p_R q_s - p^2 p_R^2 q_s^2) \quad + (\gamma_{1,2} - \gamma_{1,1})(pq_R q_S^2 - p^2 p_R^2 q_s^2) \\ & \quad + (\gamma_{2,1} - \gamma_{1,1})(pq_R^2 q_S - p^2 p_R^2 q_s^2) \Big) \\ & = \frac{1-p}{p} \gamma_{2,2} + \frac{p - \epsilon_S}{p \epsilon_S} \gamma_{2,1} + \frac{p - \epsilon_R}{p \epsilon_R} \gamma_{1,2} + \frac{(p - \epsilon_S)(p - \epsilon_R)}{p \epsilon_S \epsilon_R} \gamma_{1,1} \end{aligned}$$

A.3 Derivation of Expectation of Ψ for $\sigma' = Id$

There is equal chance for the two arrival order of i, j.

$$E[X_{i,j}] = \frac{p(q_R + q_S)}{2} I[i.k = j.k] = \frac{\varepsilon_R + \varepsilon_S}{2} I[i.k = j.k]$$

In particular, it is important to note that according to Definition 2, stream joins need to take into account timestamps, we cannot assume that a join can be generated as long as one of i, j is sampled and use the principle of inclusion-exclusion to get the following result.

$$E[X_{i,j}] = p(q_R + q_S - q_R q_S)) I[i.k = j.k]$$

The complete derivation is given by:

$$E[|\Psi|] = \sum E[X_{i,j}] = \sum_{i.k=j.k} E[X_{i,j}] = \frac{\varepsilon_R + \varepsilon_S}{2} \gamma_{1,1} = \frac{\varepsilon_R + \varepsilon_S}{2} J$$

A.4 Proof of Variance of $\hat{J}$ for $\sigma' = Id$

We still use the notations in Appendix A.3. Obviously,

$$\mathrm{Var}[\hat{J}] = 4\mathrm{Var}[|\Psi|]/(\varepsilon_R + \varepsilon_S)^2$$

and we can obtain the result by calculating $\mathrm{Var}[|\Psi|]$.

$$\begin{aligned} \mathrm{Var}[|\Psi|] &= \sum \mathrm{Cov}(X_{r,s}, X_{r',s'}) \\ &= \sum E[X_{r,s}, X_{r',s'}] - E[X_{r,s}]E[X_{r',s'}] \end{aligned}$$

There is a case where there are two tuples a, b with the same key in R, S respectively. When a is sampled by σ_R and b is not sampled by σ_S, if a comes before b then a join is generated otherwise not. So, we should discuss about $E[X_{r,s}, X_{r',s'}]$ with taking timestamps into account, which means the permutation of the arrival.

$$\begin{aligned} & E[X_{r,s}, X_{r',s'}](r \neq r', s \neq s', r.k \neq r'.k) \\ & \quad = E[X_{r,s}]E[X_{r',s'}] \\ & E[X_{r,s}, X_{r',s'}](r \neq r', s \neq s', r.k = r'.k) \\ & \quad = \frac{1}{4} p(q_R + q_S)^2 I[r.k = s.k = r'.k = s'.k] \\ & E[X_{r,s}, X_{r,s'}](s \neq s') \\ & \quad = \frac{1}{3} p(q_R + q_S q_R + q_S^2) I[r.k = s.k = s'.k] \\ & E[X_{r,s}, X_{r',s}](r \neq r') \\ & \quad = \frac{1}{3} p(q_R^2 + q_S q_R + q_S) I[r.k = s.k = r'.k] \\ & E[X_{r,s}, X_{r,s}] = E[X_{r,s}] \end{aligned}$$

So the final variance is given by:

$$\begin{aligned} & \mathrm{Var}[\hat{J}] \\ & \quad = \quad 4\mathrm{Var}[|\Psi|]/(\varepsilon_R + \varepsilon_S)^2 \\ & \quad = \quad \frac{4}{(\varepsilon_R + \varepsilon_S)^2} \\ & \qquad \Big((\gamma_{2,2} - \gamma_{2,1} - \gamma_{1,2} + \gamma_{1,1}) \\ & \qquad \times \left(\frac{1}{4} p(q_R + q_S)^2 - \frac{1}{4} p^2 (q_R + q_S)^2 \right) \\ & \qquad + (\gamma_{1,2} - \gamma_{1,1}) \left(\frac{1}{3} p(q_R + q_S q_R + q_S^2) - \frac{1}{4} p^2 (q_R + q_S)^2 \right) \end{aligned}$$

$$
\begin{aligned}
&+(\gamma_{2,1}-\gamma_{1,1})\left(\frac{1}{3}p(q_R^2+q_Sq_R+q_S)-\frac{1}{4}p^2(q_R+q_S)^2\right)\\
&+\gamma_{1,1}\left(\frac{1}{2}p(q_R+q_S)-\frac{1}{4}p^2(q_R+q_S)^2\right)\Bigg)\\
= \quad &\left(\frac{1-p}{p}\right)\gamma_{2,2}+\frac{(\varepsilon_S-3\varepsilon_R)(\varepsilon_S+\varepsilon_R)+4p\varepsilon_R}{3p(\varepsilon_R+\varepsilon_S)^2}\gamma_{2,1}\\
&+\frac{(\varepsilon_R-3\varepsilon_S)(\varepsilon_R+\varepsilon_S)+4p\varepsilon_S}{3p(\varepsilon_R+\varepsilon_S)^2}\gamma_{1,2}+\frac{-\varepsilon_R-\varepsilon_S-2p}{3p(\varepsilon_R+\varepsilon_S)}\gamma_{1,1}
\end{aligned}
$$

A.5 Derivation of Expectation of Ψ for $\sigma = RS, \sigma' = Id \circ \sigma$

For tuple t we introduce a virtual property o representing the order in which the tuple reaches the server in the two streams R, S, taking unique values from 1 to n (i.e., $|R|+|S|$). It is virtual because it can be indirectly deduced from the timestamp. For given two orders, the expectation of an indicator is as follows:

$$
E[X_{r,s}|r.o,s.o] = \begin{cases} 1 & \max(r.o,s.o) \leqslant \alpha \\ \frac{\binom{\max(r.o,s.o)-2}{\alpha-2}}{\binom{\max(r.o,s.o)}{\alpha}} & \max(r.o,s.o) > \alpha \end{cases}
$$

With equal probability of arrival order, the indicator for any two tuples has the following expectation:

$$
\begin{aligned}
E[X_{r,s}] &= \frac{I[r.k=s.k]}{(n)(n-1)}\sum_{r.o,s.o}E[X_{r,s}|r.o,s.o]\\
&= \frac{I[r.k=s.k]}{(n)(n-1)}\left(\sum_{i=2}^{\alpha}2(i-1)+\sum_{i=\alpha+1}^{n}2(i-1)\frac{\alpha(\alpha-1)}{i(i-1)}\right)\\
&= \frac{I[r.k=s.k]\alpha(\alpha-1)(1+2H_n-2H_\alpha)}{(n)(n-1)}
\end{aligned}
$$

Thus, the estimation is given by:

$$
E[\Psi] = \frac{\alpha(\alpha-1)(1+2H_n-2H_\alpha)}{(n)(n-1)}J
$$

A.6 Variance of $\hat{J}$ for $\sigma = RS, \sigma' = Id \circ \sigma$

For the sake of simplicity, we only consider the following expectation to produce join results when selected, without using a logical judgment function to indicate whether the key is the same. By taking full advantage of the property of reservoir sampling to sample a combination with equal probability at any moment after filling the reservoir, we first discuss the intermediate results with order effects.

$$
E[X_{r,s},X_{r',s}|r.o>r'.o>s.o] = \begin{cases} 1 & r.o \leqslant \alpha \\ \frac{\alpha(\alpha-1)}{r.o(r.o-1)} & r.o>\alpha, r'.o\leqslant\alpha \\ \frac{\alpha(\alpha-1)^2}{r.o(r.o-1)(r'.o-1)} & r'.o>\alpha \end{cases}
$$

$$
E[X_{r,s},X_{r',s}|r.o>s.o>r'.o] = \begin{cases} 1 & r.o \leqslant \alpha \\ \frac{\alpha(\alpha-1)}{r.o(r.o-1)} & r.o>\alpha, s.o\leqslant\alpha \\ \frac{\alpha(\alpha-1)^2}{r.o(r.o-1)(s.o-1)} & s.o>\alpha \end{cases}
$$

$$
E[X_{r,s},X_{r',s}|s.o>r.o>r'.o] = \begin{cases} 1 & s.o \leqslant \alpha \\ \frac{\binom{s.o-3}{\alpha-3}}{\binom{s.o}{\alpha}} & s.o>\alpha \end{cases}
$$

$$
E[X_{r,s},X_{r',s'}|r.o>r'.o>s.o>s'.o] = \begin{cases} 1 & r.o \leqslant \alpha \\ \frac{\alpha(\alpha-1)}{r.o(r.o-1)} & r.o>\alpha, s.o\leqslant\alpha \\ \frac{\alpha^2(\alpha-1)^2}{r.o(r.o-1)s.o(s.o-1)} & s.o>\alpha \end{cases}
$$

$$
E[X_{r,s},X_{r',s'}|r.o>s.o>r'.o>s'.o] = \begin{cases} 1 & r.o \leqslant \alpha \\ \frac{\alpha(\alpha-1)}{r.o(r.o-1)} & r.o>\alpha, s.o\leqslant\alpha \\ \frac{\alpha(\alpha-1)^2(\alpha-2)}{r.o(r.o-1)(s.o-1)(s.o-2)} & s.o>\alpha \end{cases}
$$

$$
E[X_{r,s},X_{r',s'}|r.o>s.o>s'.o>r'.o] = \begin{cases} 1 & r.o \leqslant \alpha \\ \frac{\alpha(\alpha-1)}{r.o(r.o-1)} & r.o>\alpha, s.o\leqslant\alpha \\ \frac{\alpha(\alpha-1)^2(\alpha-2)}{r.o(r.o-1)(s.o-1)(s.o-2)} & s.o>\alpha \end{cases}
$$

With equal probability of arrival order, the indicator for tuples' join has the following expectation:

$$
\begin{aligned}
&E[X_{r,s}]\\
&= \frac{I[r.k=s.k]}{(|R|+|S|)(|R|+|S|-1)}\sum_{r.o,s.o}E[X_{r,s}|r.o,s.o]\\
&= \frac{I[r.k=s.k]}{(|R|+|S|)(|R|+|S|-1)}\left(\sum_{i=2}^{\alpha}2(i-1)+\sum_{i=\alpha+1}^{|R|+|S|}2(i-1)\frac{\alpha(\alpha-1)}{i(i-1)}\right)\\
&= \frac{I[r.k=s.k]\alpha(\alpha-1)(1+2H_{|R|+|S|}-2H_\alpha)}{(|R|+|S|)(|R|+|S|-1)}
\end{aligned}
$$

There is a case where there are two tuples a, b with the same key in R, S respectively. If a comes before b, a and b are joined if and only if a is sampled by σ_R, b is sampled by σ_S, and a is still in the reservoir when b arrives. So, to calculate the variance of $\hat{J}$, we should discuss about $E[X_{r,s}, X_{r',s'}]$ with taking timestamps into account, which means the permutation of the arrival.

In case that the matches occur on one tuple in S and two different tuples in R.

$$
\begin{aligned}
&E[X_{r,s},X_{r',s}](r\neq r')\\
&= \frac{I[r.k=s.k=r'.k]}{(n)(n-1)(n-2)}\sum_{r.o,r'.o,s.o}E[X_{r,s},X_{r',s}|r.o,r'.o,s.o]\\
&= \frac{I[r.k=s.k=r'.k]}{(n)(n-1)(n-2)}\sum_{r.o,r'.o,s.o}2(E[X_{r,s},X_{r',s}|r.o>r'.o>s.o]+\\
&\quad E[X_{r,s},X_{r',s}|r.o>s.o>r'.o]+E[X_{r,s},X_{r',s}|s.o>r.o>r'.o])\\
&= \frac{I[r.k=s.k=r'.k]}{(n)(n-1)(n-2)}\sum_{r.o,r'.o,s.o}2(2E[X_{r,s},X_{r',s}|r.o>r'.o>s.o]+\\
&\quad E[X_{r,s},X_{r',s}|s.o>r.o>r'.o])\\
&= \frac{I[r.k=s.k=r'.k]}{(n)(n-1)(n-2)}((\sum_{i=3}^{\alpha}2(i-1)(i-2)+\sum_{i=\alpha+1}^{n}2\alpha(\alpha-1)\frac{\alpha(\alpha-1)}{i(i-1)}
\end{aligned}
$$

$$+\sum_{i=\alpha+2}^{n}\sum_{j=\alpha+1}^{i-1} 4(j-1)\frac{\alpha(\alpha-1)^2}{i(i-1)(j-1)}) + (\sum_{i=3}^{\alpha}(i-1)(i-2) +$$
$$\sum_{i=\alpha+1}^{n}(i-1)(i-2)\frac{\alpha(\alpha-1)(\alpha-2)}{i(i-1)(i-2)}))$$
$$= \frac{I[r.k = s.k = r'.k]}{(n)(n-1)(n-2)}((\frac{2\alpha(\alpha-1)(\alpha-2)}{3} + \frac{2\alpha(\alpha-1)^2(n-\alpha)}{n} +$$
$$4\alpha(\alpha-1)^2\left(\frac{\alpha}{n} + H_n - H_\alpha - 1\right)) + (\frac{\alpha(\alpha-1)(\alpha-2)}{3}$$
$$+\alpha(\alpha-1)(\alpha-2)(H_n - H_\alpha)))$$
$$= \frac{I[r.k = s.k = r'.k]}{(n)(n-1)(n-2)}(\alpha(\alpha-1)(\alpha-2)(1 + H_n - H_\alpha) + 2\alpha(\alpha-1)^2$$
$$\left(2H_n - 2H_\alpha + \frac{\alpha}{n} - 1\right))$$

In case that the matches occur on two different tuples in S and two different tuples in R.

$$E[X_{r,s}, X_{r',s'}](r \neq r', s \neq s')$$
$$= \frac{I[r.k = r'.k, s.k = s'.k]}{(n)(n-1)(n-2)(n-3)}\sum_{r.o,r'.o,s.o,s'.o} 8(E[X_{r,s}, X_{r',s'}|r.o > s.o$$
$$> s'.o > r'.o] + E[X_{r,s}, X_{r',s'}|r.o > s.o > r'.o > s'.o] + E[X_{r,s},$$
$$X_{r',s'}|r.o > r'.o > s.o > s'.o])$$
$$= \frac{I[r.k = r'.k, s.k = s'.k]}{(n)(n-1)(n-2)(n-3)}\sum_{r.o,r'.o,s.o,s'.o} 8(E[X_{r,s}, X_{r',s'}|r.o > r'.o$$
$$> s.o > s'.o] + 2E[X_{r,s}, X_{r',s'}|r.o > s.o > r'.o > s'.o])$$
$$= \frac{I[r.k = r'.k, s.k = s'.k]}{(n)(n-1)(n-2)(n-3)}\sum_{r.o,r'.o,s.o,s'.o} 8(($$
$$\frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)}{24} + \sum_{i=\alpha+1}^{n}(\frac{\alpha(\alpha-1)(\alpha-2)}{6} + (i-\alpha-1)$$
$$\frac{\alpha(\alpha-1)}{2})\frac{\alpha(\alpha-1)}{i(i-1)} + \sum_{i=\alpha+3}^{n}\sum_{j=\alpha+1}^{i-2}(i-j-1)(j-1)\frac{\alpha^2(\alpha-1)^2}{i(i-1)j(j-1)}))$$
$$+2(\frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)}{24} + \sum_{i=\alpha+1}^{n}\left(\frac{\alpha(\alpha-1)(\alpha-2)}{6}\right)\frac{\alpha(\alpha-1)}{i(i-1)}$$
$$+\sum_{i=\alpha+2}^{n}\sum_{j=\alpha+1}^{i-1}\frac{(j-1)(j-2)}{2}\frac{\alpha(\alpha-1)^2(\alpha-2)}{i(i-1)(j-1)(j-2)})))$$
$$= \frac{I[r.k = r'.k, s.k = s'.k]}{(n)(n-1)(n-2)(n-3)}(\alpha(\alpha-1)(\alpha-2)(\alpha-3) +$$
$$4\alpha^2(\alpha-1)^2(\alpha-2)\sum_{i=\alpha+1}^{n}\frac{1}{i(i-1)} + 4\alpha^2(\alpha-1)^2\sum_{i=\alpha+1}^{n}\frac{i-1-\alpha}{i(i-1)} +$$
$$8\alpha^2(\alpha-1)^2\sum_{i=\alpha+3}^{n}\sum_{j=\alpha+1}^{i-2}\frac{i-j+1}{i(i-1)j} + 8\alpha(\alpha-1)^2(\alpha-2)\sum_{i=\alpha+2}^{n}$$
$$\sum_{j=\alpha+1}^{i-1}\frac{1}{i(i-1)})$$
$$= \frac{I[r.k = r'.k, s.k = s'.k]}{(n)(n-1)(n-2)(n-3)}(\alpha(\alpha-1)(\alpha-2)(\alpha-3) +$$
$$4\alpha^2(\alpha-1)^2(\alpha-2)(\frac{1}{\alpha} - \frac{1}{n}) +$$
$$8\alpha(\alpha-1)^2(\alpha-2)(H_n - H_{\alpha+1} - \frac{\alpha}{\alpha+1} + \frac{\alpha}{n}) +$$
$$4\alpha^2(\alpha-1)^2(H_n^2 - H_n^{(2)} - H_{\alpha+2}^2 + H_{\alpha+2}^{(2)} - 2H_nH_\alpha$$
$$+2H_\alpha H_{\alpha+2} - H_n + H_\alpha - \frac{2}{\alpha+2} - \frac{\alpha}{n} + 1 + \frac{2}{\alpha+1}))$$

In case that the matches occur on one tuple in S and one tuple in R.

$$E[X_{r,s}, X_{r,s}] = E[X_{r,s}]$$

On the basis of considering the probability, combining all the above cases, the final variance of $|\Psi|$ is given by:

$$Var[|\Psi|]$$
$$= \sum Cov(X_{r,s}, X_{r',s'})$$
$$= \sum E[X_{r,s}, X_{r',s'}] - E[X_{r,s}]E[X_{r',s'}]$$
$$= \frac{\gamma_{1,1}}{n(n-1)}\alpha(\alpha-1)(1 + 2H_n - 2H_\alpha)$$
$$+\frac{(\gamma_{1,2} - \gamma_{1,1}) + (\gamma_{2,1} - \gamma_{1,1})}{n(n-1)(n-2)}(\alpha(\alpha-1)(\alpha-2)$$
$$(1 + H_n - H_\alpha) + 2\alpha(\alpha-1)^2(2H_n - 2H_\alpha + \frac{\alpha}{n} - 1))$$
$$+\frac{I[r.k = s.k, r'.k = s'.k]}{n(n-1)(n-2)(n-3)}(\alpha(\alpha-1)(\alpha-2)(\alpha-3)$$
$$+4\alpha^2(\alpha-1)^2(\alpha-2)(\frac{1}{\alpha} - \frac{1}{n}) + 8\alpha(\alpha-1)^2(\alpha-2)$$
$$(H_n - H_{\alpha+1} - \frac{\alpha}{\alpha+1} + \frac{\alpha}{n})) + 4\alpha^2(\alpha-1)^2$$
$$(H_n^2 - H_n^{(2)} - H_{\alpha+2}^2 + H_{\alpha+2}^{(2)} - 2H_nH_\alpha + 2H_\alpha H_{\alpha+2}$$
$$-H_n + H\alpha - \frac{2}{\alpha+2} - \frac{\alpha}{n} + 1 + \frac{2}{\alpha+1}))$$
$$-\frac{I[r.k = s.k]I[r'.k = s'.k]}{n^2(n-1)^2}(1 + 2H_n - 2H_\alpha)^2\alpha^2(\alpha-1)^2$$
$$= \gamma_{1,1}(\frac{\alpha(\alpha-1)}{(n-1)(n-2)}(1 + 2H_n - 2H_\alpha) + \frac{\alpha(\alpha-1)}{n(n-1)(n-2)}$$
$$((-10\alpha + 8)H_n + (10\alpha - 8)H_\alpha + 2\alpha - 2 - \frac{4\alpha(\alpha-1)}{n})$$
$$+\frac{1}{n(n-1)(n-2)(n-3)}(\alpha(\alpha-1)(\alpha-2)(\alpha-3)$$
$$+4\alpha^2(\alpha-1)^2(\alpha-2)(\frac{1}{\alpha} - \frac{1}{n}) + 8\alpha(\alpha-1)^2(\alpha-2)$$
$$(H_n - H_{\alpha+1} - \frac{\alpha}{\alpha+1} + \frac{\alpha}{n})) + \frac{\alpha^2(\alpha-1)^2}{(n-1)^2(n-2)(n-3)}$$
$$(-4H_\alpha^2 - 4H_{\alpha+2}^2 - 4H_n^{(2)} + 4H_{\alpha+2}^{(2)} - 8H_n + 8H_\alpha + 3$$
$$+8H_\alpha H_{\alpha+2} - \frac{4\alpha}{n} - \frac{8}{\alpha+2} + \frac{8}{\alpha+1})$$
$$+\frac{\alpha^2(\alpha-1)^2}{n(n-1)^2(n-2)(n-3)}(16H_n^2 + 20H_\alpha^2 + 4H_n^{(2)}$$
$$+4H_{\alpha+2}^2 - 4H_{\alpha+2}^{(2)} - 32H_nH_\alpha - 8H_\alpha H_{\alpha+2} + 24H_n$$
$$-24H_\alpha + 1 + \frac{4\alpha}{n} + \frac{8}{\alpha+2} - \frac{8}{\alpha+1})$$
$$+\frac{6\alpha^2(\alpha-1)^2}{n^2(n-1)^2(n-2)(n-3)}(1 + 4H_n^2 + 4H_\alpha^2 + 4H_n$$
$$-4H_\alpha - 8H_nH_\alpha))$$
$$+\gamma_{2,2}(\frac{1}{n(n-1)(n-2)(n-3)}(\alpha(\alpha-1)(\alpha-2)(\alpha-3)$$

$$+4\alpha(\alpha-1)^2(\alpha-2)(1+\frac{\alpha}{n}+2H_n-2H_{\alpha+1}-\frac{2\alpha}{\alpha+1}))$$
$$+\frac{\alpha^2(\alpha-1)^2}{(n-1)^2(n-2)(n-3)}(-4H_\alpha^2-4H_{\alpha+2}^2-4H_n^{(2)}+4H_{\alpha+2}^{(2)}$$
$$-8H_n+8H_\alpha+3+8H_\alpha H_{\alpha+2}-\frac{4\alpha}{n}$$
$$-\frac{8}{\alpha+2}+\frac{8}{\alpha+1})+\frac{\alpha^2(\alpha-1)^2}{n(n-1)^2(n-2)(n-3)}(16H_n^2+20H_\alpha^2$$
$$+4H_n^{(2)}+4H_{\alpha+2}^2-4H_{\alpha+2}^{(2)}-32H_nH_\alpha-8H_\alpha H_{\alpha+2}+24H_n$$
$$-24H_\alpha+1+\frac{4\alpha}{n}+\frac{8}{\alpha+2}-\frac{8}{\alpha+1})+\frac{6\alpha^2(\alpha-1)^2}{n^2(n-1)^2(n-2)(n-3)}$$
$$(1+4H_n^2+4H_\alpha^2+4H_n-4H_\alpha-8H_nH_\alpha))$$
$$+(\gamma_{1,2}+\gamma_{2,1})(\frac{1}{n(n-1)(n-2)}(\alpha(\alpha-1)(\alpha-2)(1+H_n-H_\alpha)$$
$$+2\alpha(\alpha-1)^2(2H_n-2H_\alpha+\frac{\alpha}{n}-1))$$
$$-\frac{1}{n(n-1)(n-2)(n-3)}(\alpha(\alpha-1)(\alpha-2)(\alpha-3)+4\alpha(\alpha-1)^2$$
$$(\alpha-2)(1+\frac{\alpha}{n}+2H_n-2H_{\alpha+1}-\frac{2\alpha}{\alpha+1}))+\frac{\alpha^2(\alpha-1)^2}{(n-1)^2(n-2)(n-3)}$$
$$(-4H_\alpha^2-4H_{\alpha+2}^2-4H_n^{(2)}+4H_{\alpha+2}^{(2)}-8H_n+8H_\alpha+3+8H_\alpha H_{\alpha+2}$$
$$-\frac{4\alpha}{n}-\frac{8}{\alpha+2}+\frac{8}{\alpha+1})+\frac{\alpha^2(\alpha-1)^2}{n(n-1)^2(n-2)(n-3)}(16H_n^2+20H_\alpha^2$$
$$+4H_n^{(2)}+4H_{\alpha+2}^2-4H_{\alpha+2}^{(2)}-32H_nH_\alpha-8H_\alpha H_{\alpha+2}+24H_n$$
$$-24H_\alpha+1+\frac{4\alpha}{n}+\frac{8}{\alpha+2}-\frac{8}{\alpha+1})+\frac{6\alpha^2(\alpha-1)^2}{n^2(n-1)^2(n-2)(n-3)}$$
$$(1+4H_n^2+4H_\alpha^2+4H_n-4H_\alpha-8H_nH_\alpha))$$
$$=\gamma_{1,1}\Theta(\frac{\alpha^2\log\frac{n}{\alpha}}{n^2})+\gamma_{2,2}\Theta(\frac{\alpha^4\log\frac{n}{\alpha}}{n^4})$$
$$+(\gamma_{1,2}+\gamma_{2,1})\Theta(\frac{\alpha^3\log\frac{n}{\alpha}}{n^3})$$

The final variance of $\hat{J}$ is given by:

$$\begin{aligned}Var[\hat{J}] &= \frac{n^2(n-1)^2}{\alpha^2(\alpha-1)^2(1+2H_n-2H_\alpha)^2}Var[\Psi]\\ &= \gamma_{1,1}(\frac{n^2}{\alpha^2\log\frac{n}{\alpha}})+\gamma_{2,2}(\frac{1}{\log\frac{n}{\alpha}})+(\gamma_{1,2}+\gamma_{2,1})\Theta(\frac{n}{\alpha\log\frac{n}{\alpha}})\end{aligned}$$

A.7 Derivation of Expectation of Ψ for $\sigma = RS, \sigma' = Id$

For given two orders, the expectation of an indicator is as follows:

$$E[X_{r,s}|r.o,s.o]=\begin{cases}1 & \max(r.o,s.o)\leqslant\alpha\\ \frac{\alpha}{\max(r.o,s.o)} & \max(r.o,s.o)>\alpha\end{cases}$$

With equal probability of arrival order, the indicator for tuples' join has the following expectation:

$$\begin{aligned}E[X_{r,s}] &= 2\frac{I[r.k=s.k]}{n(n-1)}(\frac{\alpha(\alpha-1)}{2}+\sum_{i=\alpha+1}^{n}(i+1)\frac{\alpha}{i})\\ &= \frac{I[r.k=s.k]}{n(n-1)}(2\alpha n-2\alpha H_\alpha-\alpha(\alpha+1)+2\alpha H_\alpha)\end{aligned}$$

Thus, the estimation is given by:

$$E[\Psi]=\frac{2\alpha n-2\alpha H_\alpha-\alpha(\alpha+1)+2\alpha H_\alpha}{n(n-1)}J$$

A.8 Variance of $\hat{J}$ for $\sigma = RS, \sigma' = Id$

Just like the condition of $\sigma = \sigma' = RS$, we also only consider the following expectation to produce join results when selected, without using a logical judgment function to indicate whether the key is the same. By taking full advantage of the property of reservoir sampling to sample a combination with equal probability at any moment after filling the reservoir, we first discuss the intermediate results with order effects.

$$E[X_{r,s},X_{r',s}|r.o>r'.o>s.o]=\begin{cases}1 & r.o\leqslant\alpha\\ \frac{\alpha}{r.o} & r.o>\alpha, r'.o\leqslant\alpha\\ \frac{\alpha}{r.o} & r'.o>\alpha\end{cases}$$

$$E[X_{r,s},X_{r',s}|s.o>r.o>r'.o]=\begin{cases}1 & s.o\leqslant\alpha\\ \frac{\alpha(\alpha-1)}{s.o(s.o-1)} & s.o>\alpha\end{cases}$$

$$E[X_{r,s},X_{r',s'}|r.o>s.o>>r'.o>s'.o]=\begin{cases}1 & r.o\leqslant\alpha\\ \frac{\alpha}{r.o} & r.o>\alpha, r'.o\leqslant\alpha\\ \frac{\alpha^2}{r.or'.o} & r'.o>\alpha\end{cases}$$

$$E[X_{r,s},X_{r',s'}|r.o>r'.o>s.o>s'.o]=\begin{cases}1 & r.o\leqslant\alpha\\ \frac{\alpha}{r.o} & r.o>\alpha, r'.o\leqslant\alpha\\ \frac{\alpha(\alpha-1)}{r.o(r'.o-1)} & r'.o>\alpha\end{cases}$$

With equal probability of arrival order, the indicator for tuples' join has the following expectation:

$$\begin{aligned}&E[X_{r,s}]\\ &= 2\frac{I[r.k=s.k]}{n(n-1)}(\frac{\alpha(\alpha-1)}{2}+\sum_{i=\alpha+1}^{n}(i+1)\frac{\alpha}{i})\\ &= \frac{I[r.k=s.k]}{n(n-1)}(2\alpha n-2\alpha H_n-\alpha(\alpha+1)+2\alpha H_\alpha)\end{aligned}$$

There is a case where there are two tuples a, b with the same key in R, S respectively. If a comes before b, a and b are joined if and only if a is sampled by σ_R and a is still in the reservoir when b arrives. So, to calculate the variance of $\hat{J}$, we should discuss about $E[X_{r,s}, X_{r',s'}]$ with taking timestamps into account, which means the permutation of the arrival.

In case that the matches occur on one tuple in S and two different tuples in R.

$$\begin{aligned}&E[X_{r,s}X_{r',s}])(r\neq r')\\ &= \frac{I[r.k=r'.k=s.k]}{n(n-1)(n-2)}\sum_{r.o,r'.o,s.o}2(EX_{[r,s}X_{r',s}|r.o>r'.o>s.o]\\ &\quad +EX_{[r,s}X_{r',s}|r.o>s.o>r'.o]+EX_{[r,s}X_{r',s}|s.o>r.o>r'.o])\end{aligned}$$

$$
\begin{aligned}
&= \frac{I[r.k = r'.k = s.k]}{n(n-1)(n-2)} \sum_{r.o, r'.o, s.o} (4EX_{[r,s}X_{r',s}|r.o > r'.o > s.o] \\
&\quad +2EX_{[r,s}X_{r',s}|s.o > r.o > r'.o]) \\
&= \frac{I[r.k = r'.k = s.k]}{n(n-1)(n-2)} (4(\sum_{i=3}^{\alpha} \frac{(i-1)(i-2)}{2} + \sum_{i=\alpha+1}^{n} \frac{\alpha(\alpha-1)}{2} \cdot \frac{\alpha}{i} \\
&\quad + \sum_{i=\alpha+2}^{n} \sum_{j=\alpha+1}^{i-1} (j-1)\frac{\alpha}{i}) + 2(\sum_{i=3}^{n} \frac{(i-1)(i-2)}{2} \\
&\quad + \sum_{i=\alpha+1}^{n} (i-1)(i-2)\frac{\alpha(\alpha-1)}{i(j-1)})) \\
&= \frac{I[r.k = r'.k = s.k]}{n(n-1)(n-2)} (\alpha n^2 + \alpha(2\alpha-7)n + (2\alpha^3 - 7\alpha^2 + 5\alpha + 2)H_n \\
&\quad -2\alpha(\alpha-1)(\alpha-2)H_\alpha + (\alpha-2)(\alpha+1)H_{\alpha+1} - 2\alpha(\alpha^2 - \alpha - 3))
\end{aligned}
$$

In case that the matches occur on two different tuples in S and two different tuples in R.

$$
\begin{aligned}
&E[X_{r,s}X_{r',s'}](r \neq r, s \neq s') \\
&= \frac{I[r.k = s.k, r'.k = s'.k]}{n(n-1)(n-2)(n-3)} (8E[X_{r,s}X_{r',s'}|r.o > s.o > r'.o > s'.o] \\
&\quad +16E[X_{r,s}X_{r',s'}|r.o > r'.o > s.o > s'.o]) \\
&= \frac{I[r.k = s.k, r'.k = s'.k]}{n(n-1)(n-2)(n-3)} (8(\frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)}{24} \\
&\quad + \sum_{i=\alpha+1}^{n} (\frac{\alpha(\alpha-1)(\alpha-2)}{6} + \frac{(i-\alpha-1)\alpha(\alpha-1)}{2})\frac{\alpha}{i} \\
&\quad + \sum_{i=\alpha+3}^{n} \sum_{j=\alpha+1}^{i-2} (i-j-1)(j-1)\frac{\alpha(\alpha-1)}{i(j-1)}) \\
&\quad +16(\frac{\alpha(\alpha-1)(\alpha-2)(\alpha-3)}{24} + \sum_{i=\alpha+1}^{n} \frac{\alpha(\alpha-1)(\alpha-2)}{6} \cdot \frac{\alpha}{i} \\
&\quad + \sum_{i=\alpha+2}^{n} \sum_{j=\alpha+1}^{i-1} \frac{(j-1)(j-2)}{2} \cdot \frac{\alpha(\alpha-1)}{i(j-1)})) \\
&= \frac{I[r.k = s.k, r'.k = s'.k]}{n(n-1)(n-2)(n-3)} (4\alpha(\alpha-1)n^2 - 4\alpha(\alpha-1)(\alpha+1)n \\
&\quad +\alpha(\alpha-1)(\alpha^2 + 3\alpha - 14) - 12\alpha(\alpha-1)(\alpha-2)H_n + 12\alpha^2(\alpha-1) \\
&\quad H_\alpha - 4\alpha(\alpha-1)^2(\alpha-2)H_{\alpha+2} + 4\alpha(\alpha-1)(\alpha+1)(\alpha-4)H_{\alpha+1})
\end{aligned}
$$

On the basis of considering the probability, combining all the above cases, the final variance of $|\Psi|$ is given by:

$$
\begin{aligned}
&Var[|\Psi|] \\
&= \sum Cov(X_{r,s}, X_{r',s'}) \\
&= \sum E[X_{r,s}, X_{r',s'}] - E[X_{r,s}]E[X_{r',s'}] \\
&= \frac{\gamma_{1,1}}{n(n-1)}(2\alpha n - 2\alpha H_n - \alpha(\alpha+1) + 2\alpha H_\alpha) \\
&\quad + \frac{\gamma_{1,2} + \gamma_{2,1}}{n(n-1)(n-2)} (\alpha n^2 + \alpha(2\alpha-7)n + (2\alpha^3 - 7\alpha^2 + 5\alpha + 2)H_n \\
&\quad -2\alpha(\alpha-1)(\alpha-2)H_\alpha + (\alpha-2)(\alpha+1)H_{\alpha+1} - 2\alpha(\alpha^2 - \alpha - 3)) \\
&\quad + \frac{\gamma_{1,1} + \gamma_{2,2} - \gamma_{1,2} - \gamma_{2,1}}{n(n-1)(n-2)(n-3)} (4\alpha(\alpha-1)n^2 - 4\alpha(\alpha-1)(\alpha+1)n \\
&\quad +\alpha(\alpha-1)(\alpha^2 + 3\alpha - 14) - 12\alpha(\alpha-1)(\alpha-2)H_n + 12\alpha^2(\alpha-1) \\
&\quad H_\alpha - 4\alpha(\alpha-1)^2(\alpha-2)H_{\alpha+2} + 4\alpha(\alpha-1)(\alpha+1)(\alpha-4)H_{\alpha+1}) \\
&\quad - \frac{\gamma_{1,1} + \gamma_{2,2} - \gamma_{1,2} - \gamma_{2,1}}{n(n-1)(n-2)(n-3)} (2\alpha n - 2\alpha H_n - \alpha(\alpha+1) + 2\alpha H_\alpha)^2 \\
&= \gamma_{1,1}(\frac{-4\alpha}{(n-1)(n-2)} + \frac{\alpha(-1-\alpha+2H_\alpha)}{n(n-1)} - \frac{2\alpha H_n}{n(n-1)} \\
&\quad - \frac{2\alpha(2\alpha-7)}{(n-1)(n-2)} \quad - \frac{2}{n(n-1)(n-2)}(-2\alpha(\alpha^2 - \alpha - 3) \\
&\quad +(2\alpha^3 - 7\alpha^2 + 5\alpha + 2)H_n - 2\alpha(\alpha-1)(\alpha-2)H_\alpha + (\alpha-2) \\
&\quad (\alpha+1)H_{\alpha+1}) - \frac{4\alpha n^2 - 4\alpha(4\alpha+3)n}{(n-1)^2(n-2)(n-3)} + \frac{1}{(n-1)^2(n-2)(n-3)} \\
&\quad (-2\alpha(8\alpha^2 + 19\alpha - 5) + 16\alpha^2(\alpha+2)H_\alpha + 4\alpha(\alpha-1)(\alpha+1)(\alpha-4) \\
&\quad H_{\alpha+1} - 4\alpha(\alpha-1)^2(\alpha-2)H_{\alpha+2} - 4\alpha^2 H_\alpha^2) \\
&\quad + \frac{1}{n(n-1)^2(n-2)(n-3)}(2\alpha(2\alpha^3 + 4\alpha^2 + 11\alpha - 7) - 8\alpha^2 \\
&\quad (4\alpha+1)H_\alpha + 4\alpha(\alpha-1)^2(\alpha-1)H_{\alpha+2} - 4\alpha(\alpha-1)(\alpha+1) \\
&\quad (\alpha-4)H_{\alpha+1} + 20\alpha^2 H_{\alpha^2}) - \frac{6}{n^2(n-1)^2(n-2)(n-3)} \\
&\quad (\alpha^2(\alpha+1)^2 + 4\alpha^2 H_\alpha^2 - 4\alpha^2(\alpha+1)H_\alpha) \\
&\quad + \frac{H_n}{n^2(n-1)^2(n-2)(n-3)}(8\alpha^2 n^3 - 8\alpha n^2(2\alpha^2 - 4\alpha + 3 - \alpha H_\alpha) \\
&\quad +8\alpha n(4\alpha^2 + 4\alpha + 3 - 5\alpha H_\alpha) - 24\alpha^2(\alpha + 1 - 2H_\alpha)) - \frac{4\alpha^2 H_n^2}{n^2(n-1)^2}) \\
&\quad (\gamma_{1,2} + \gamma_{2,1})(\frac{\alpha n}{(n-1)(n-2)} + \frac{\alpha(2\alpha-7)}{(n-1)(n-2)} \\
&\quad + \frac{1}{n(n-1)(n-2)}(-2\alpha(\alpha^2 - \alpha - 3) \\
&\quad +(2\alpha^3 - 7\alpha^2 + 5\alpha + 2)H_n - 2\alpha(\alpha-1)(\alpha-2)H_\alpha + (\alpha-2) \\
&\quad (\alpha+1)H_{\alpha+1}) + \frac{4\alpha n^2 - 4\alpha(4\alpha+3)n}{(n-1)^2(n-2)(n-3)} - \frac{1}{(n-1)^2(n-2)(n-3)} \\
&\quad (-2\alpha(8\alpha^2 + 19\alpha - 5) + 16\alpha^2(\alpha+2)H_\alpha + 4\alpha(\alpha-1)(\alpha+1)(\alpha-4) \\
&\quad H_{\alpha+1} - 4\alpha(\alpha-1)^2(\alpha-2)H_{\alpha+2} - 4\alpha^2 H_\alpha^2) \\
&\quad - \frac{1}{n(n-1)^2(n-2)(n-3)}(2\alpha(2\alpha^3 + 4\alpha^2 + 11\alpha - 7) - 8\alpha^2 \\
&\quad (4\alpha+1)H_\alpha + 4\alpha(\alpha-1)^2(\alpha-1)H_{\alpha+2} - 4\alpha(\alpha-1)(\alpha+1) \\
&\quad (\alpha-4)H_{\alpha+1} + 20\alpha^2 H_{\alpha^2}) + \frac{6}{n^2(n-1)^2(n-2)(n-3)} \\
&\quad (\alpha^2(\alpha+1)^2 + 4\alpha^2 H_\alpha^2 - 4\alpha^2(\alpha+1)H_\alpha) \\
&\quad - \frac{H_n}{n^2(n-1)^2(n-2)(n-3)}(8\alpha^2 n^3 - 8\alpha n^2(2\alpha^2 - 4\alpha + 3 - \alpha H_\alpha) \\
&\quad +8\alpha n(4\alpha^2 + 4\alpha + 3 - 5\alpha H_\alpha) - 24\alpha^2(\alpha + 1 - 2H_\alpha)) + \frac{4\alpha^2 H_n^2}{n^2(n-1)^2}) \\
&\quad +\gamma_{2,2}(-\frac{4\alpha n^2 - 4\alpha(4\alpha+3)n}{(n-1)^2(n-2)(n-3)} + \frac{1}{(n-1)^2(n-2)(n-3)} \\
&\quad (-2\alpha(8\alpha^2 + 19\alpha - 5) + 16\alpha^2(\alpha+2)H_\alpha + 4\alpha(\alpha-1)(\alpha+1)(\alpha-4) \\
&\quad H_{\alpha+1} - 4\alpha(\alpha-1)^2(\alpha-2)H_{\alpha+2} - 4\alpha^2 H_\alpha^2) \\
&\quad + \frac{1}{n(n-1)^2(n-2)(n-3)}(2\alpha(2\alpha^3 + 4\alpha^2 + 11\alpha - 7) - 8\alpha^2 \\
&\quad (4\alpha+1)H_\alpha + 4\alpha(\alpha-1)^2(\alpha-1)H_{\alpha+2} - 4\alpha(\alpha-1)(\alpha+1) \\
&\quad (\alpha-4)H_{\alpha+1} + 20\alpha^2 H_{\alpha^2}) - \frac{6}{n^2(n-1)^2(n-2)(n-3)}
\end{aligned}
$$

$$
\begin{aligned}
&(\alpha^2(\alpha+1)^2 + 4\alpha^2 H_\alpha^2 - 4\alpha^2(\alpha+1)H_\alpha) \\
&+\frac{H_n}{n^2(n-1)^2(n-2)(n-3)}(8\alpha^2 n^3 - 8\alpha n^2(2\alpha^2 - 4\alpha + 3 - \alpha H_\alpha) \\
&+8\alpha n(4\alpha^2 + 4\alpha + 3 - 5\alpha H_\alpha) - 24\alpha^2(\alpha + 1 - 2H_\alpha)) - \frac{4\alpha^2 H_n^2}{n^2(n-1)^2}) \\
= {}& \gamma_{1,1}\Theta(\frac{\alpha}{n} + \frac{\alpha^3 \log \frac{n}{\alpha}}{n^3}) + \gamma_{2,2}\Theta(\frac{\alpha}{n^2} + \frac{\alpha^2 \log n}{n^3} + \frac{\alpha^3 \log \frac{n}{\alpha}}{n^4}) \\
&+(\gamma_{1,2} + \gamma_{2,1})\Theta(\frac{\alpha}{n} + \frac{\alpha^3 \log \frac{n}{\alpha}}{n^3})
\end{aligned}
$$

The final variance of $\hat{J}$ is given by:

$$
\begin{aligned}
Var[\hat{J}] &= \frac{n^2(n-1)^2}{(2\alpha n - 2\alpha H_\alpha - \alpha(\alpha+1) + 2\alpha H_\alpha)^2} Var[\Psi] \\
&= \gamma_{1,1}\Theta(\frac{n}{\alpha} + \frac{\alpha \log \frac{n}{\alpha}}{n}) + \gamma_{2,2}\Theta(\frac{1}{\alpha} + \frac{\log n}{n} + \frac{\alpha \log \frac{n}{\alpha}}{n^2}) \\
&\quad +(\gamma_{1,2} + \gamma_{2,1})\Theta(\frac{n}{\alpha} + \frac{\alpha \log \frac{n}{\alpha}}{n})
\end{aligned}
$$