Pass-and-Swap Queues

Joint work with Jan-Pieter Dorsman (UvA)

Céline Comte — c.m.comte@tue.nl
Eindhoven University of Technology

2021 INFORMS Annual Meeting — Session “Recent Advances in Load Balancing”
The M/M/1 queue

Model

• Jobs arrive according to a Poisson process with rate $\lambda$.
• Service times are i.i.d. exponentially distributed with rate $\mu$.
• A single server.

Analysis

• Markov (birth-and-death) process.
• Stationary distribution: $\pi(n) = \left(1 - \frac{\lambda}{\mu}\right) \frac{\lambda^n}{n!}$.
• Proof: (partial) balance equations + normalization equation.
The M/M/1 queue

Model
- Jobs arrive according to a Poisson process with rate $\lambda$.
- Service times are i.i.d. exponentially distributed with rate $\mu$.
- A single server.

Analysis
- Markov (birth-and-death) process.
- Stationary distribution: $\pi(n) = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n$.
- Proof: (partial) balance equations + normalization equation.
Order-independent queues

- Redundancy cancel-on-complete (Gardner et al., 2016)
- Order-independent (OI) queue (Berezner et al., 1995) (Bonald and Comte, 2017)
Order-independent queues

- Redundancy cancel-on-complete (Gardner et al., 2016)

\[ \lambda_A, \lambda_B, \lambda_C, \mu_1, \mu_2, \mu_3 \]
Order-independent queues

- **Redundancy cancel-on-complete** (Gardner et al., 2016)

\[ \lambda_A \rightarrow \mu_1 \rightarrow \mu_2 \rightarrow \mu_3 \]
Order-independent queues

• Redundancy cancel-on-complete (Gardner et al., 2016)
Order-independent queues

• Redundancy cancel-on-complete (Gardner et al., 2016)
Order-independent queues

- **Redundancy cancel-on-complete** (Gardner et al., 2016)
Order-independent queues

• **Redundancy cancel-on-complete** (Gardner et al., 2016)
Order-independent queues

- **Redundancy cancel-on-complete** (Gardner et al., 2016)
Order-independent queues

- **Redundancy cancel-on-complete** (Gardner et al., 2016)
Order-independent queues

- Redundancy cancel-on-complete (Gardner et al., 2016)
Order-independent queues

- Redundancy cancel-on-complete (Gardner et al., 2016)

- Order-independent (OI) queue (Berezner et al., 1995) (Bonald and Comte, 2017)
Order-independent queues

• Product-form stationary distribution (Gardner et al., 2016):

\[
\pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}.
\]

• Proof: (partial) balance equations + normalization equation.

• Why study product-form queues?
  ▶ Rich in applications.
  ▶ Exact performance analysis is not completely hopeless.
Order-independent queues

• Product-form stationary distribution (Gardner et al., 2016):

\[ \pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}. \]

• Proof: (partial) balance equations + normalization equation.
Order-independent queues

• Product-form stationary distribution (Gardner et al., 2016):

\[ \pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}. \]

• Proof: (partial) balance equations + normalization equation.

• Why study product-form queues?
  ▶ Rich in applications.
  ▶ Exact performance analysis is not completely hopeless.
Current state of the art

• Fundamental contributions by Kelly and Whittle in the 1970’s and 1980’s.
Current state of the art

- Fundamental contributions by Kelly and Whittle in the 1970’s and 1980’s.
- Study of product-form queues in different contexts: redundancy scheduling, matching systems, manufacturing systems, etc.

Q: Can we build a model that captures all product-form queues out there?
A: Several frameworks have been constructed recently:
- Adan, Kleiner, Righter, Weiss (2018)
- Gardner, Righter (2020)
- Ayesta, Bodas, Dorsman, Verloop (2021)

A: But, still, new product-form queues keep appearing, not captured by these frameworks, such as the pass-and-swap queue.
Current state of the art

- Fundamental contributions by Kelly and Whittle in the 1970’s and 1980’s.
- Study of product-form queues in different contexts: redundancy scheduling, matching systems, manufacturing systems, etc.
- Q: Can we build a model that captures all product-form queues out there?

Several frameworks have been constructed recently:

▶ Adan, Kleiner, Righter, Weiss (2018)
▶ Gardner, Righter (2020)
▶ Ayesta, Bodas, Dorsman, Verloop (2021)

But, still, new product-form queues keep appearing, not captured by these frameworks, such as the pass-and-swap queue.
Current state of the art

• Fundamental contributions by Kelly and Whittle in the 1970’s and 1980’s.
• Study of product-form queues in different contexts: redundancy scheduling, matching systems, manufacturing systems, etc.
• Q: Can we build a model that captures all product-form queues out there?
• A: Several frameworks have been constructed recently:
  ➤ Adan, Kleiner, Righter, Weiss (2018)
  ➤ Gardner, Righter (2020)
  ➤ Ayesta, Bodas, Dorsman, Verloop (2021)
Current state of the art

• Fundamental contributions by Kelly and Whittle in the 1970’s and 1980’s.
• Study of product-form queues in different contexts: redundancy scheduling, matching systems, manufacturing systems, etc.
• Q: Can we build a model that captures all product-form queues out there?
• A: Several frameworks have been constructed recently:
  ▶ Adan, Kleiner, Righter, Weiss (2018)
  ▶ Gardner, Righter (2020)
  ▶ Ayesta, Bodas, Dorsman, Verloop (2021)
• A: But, still, new product-form queues keep appearing, not captured by these frameworks, such as the pass-and-swap queue.
Definition

Pass-and-swap (P&S) queues

• are an extension of OI queues,
• add a whole new dimension to product-form queues: intra-queue routing,
• have many applications.
Definition

Pass-and-swap (P&S) queues
- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

\[ \lambda_A \rightarrow B \quad \rightarrow \mu(\cdot) \rightarrow C \]

\[ \lambda_B \rightarrow A \rightarrow C \]

\[ \lambda_C \rightarrow B \rightarrow A \]

Swapping graph

\[ A \quad B \quad C \]
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

**OI queue**

\[
\lambda_A \rightarrow B \rightarrow A \rightarrow C \rightarrow B \rightarrow C \rightarrow A
\]

\[
\mu(\cdot)
\]

**Swapping graph**

\[
A \rightarrow B \rightarrow C
\]
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

![OI queue and Swapping graph](image)
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

![OI queue diagram]

![Swapping graph diagram]
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues
- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues
• are an extension of OI queues,
• add a whole new dimension to product-form queues: intra-queue routing,
• have many applications.

OI queue

Swapping graph

\[ \lambda_A \]

\[ \lambda_B \]

\[ \lambda_C \]
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

\[ \lambda_A \] \rightarrow C \left\{ \begin{array}{c} B \rightarrow A \rightarrow B \rightarrow A \rightarrow C \rightarrow A \end{array} \right\} \mu(\cdot) \rightarrow \text{Swapping graph}

\[ \lambda_B \] \rightarrow \text{B}

\[ \lambda_C \] \rightarrow \text{C}
Definition

Pass-and-swap (P&S) queues

• are an extension of OI queues,
• add a whole new dimension to product-form queues: intra-queue routing,
• have many applications.

OI queue

\[ \lambda_A \quad \lambda_B \quad \lambda_C \]

Swapping graph

\[ B \quad C \quad A \]
Definition

Pass-and-swap (P&S) queues
• are an extension of OI queues,
• add a whole new dimension to product-form queues: intra-queue routing,
• have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues
- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

\[ \lambda_A \]

\[ \lambda_B \]

\[ \lambda_C \]

Swapping graph

A

B

C
Definition

Pass-and-swap (P&S) queues
- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

\[
\begin{align*}
\lambda_A & \quad \rightarrow \quad C & \quad \rightarrow \quad A & \quad \rightarrow \quad B & \quad \rightarrow \quad A & \quad \rightarrow \quad C \\
\lambda_B & \quad \rightarrow \quad B & \quad \rightarrow \quad \mu(\cdot) & \quad \rightarrow \\
\lambda_C & \quad \rightarrow
\end{align*}
\]

Swapping graph

\[
A \quad \rightarrow \quad B \quad \rightarrow \quad C
\]
Definition

Pass-and-swap (P&S) queues

- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

![Diagram of OI queue and swapping graph]
Definition

Pass-and-swap (P&S) queues
- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues

• are an extension of OI queues,
• add a whole new dimension to product-form queues: intra-queue routing,
• have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues
- are an extension of OI queues,
- add a whole new dimension to product-form queues: intra-queue routing,
- have many applications.

OI queue

Swapping graph
Definition

Pass-and-swap (P&S) queues
• are an extension of OI queues,
• add a whole new dimension to product-form queues: intra-queue routing,
• have many applications.

OI queue

\[
\begin{align*}
\lambda_A & \quad \rightarrow & \quad C & \quad B & \quad A & \quad C \\
\lambda_B & \quad \rightarrow & \quad & \quad & \quad & \mu(\cdot) \\
\lambda_C & \quad \rightarrow & \quad & \quad & \quad & \rightarrow \\
\end{align*}
\]

Swapping graph

\[
\begin{align*}
A & \quad \rightarrow & \quad B & \quad \rightarrow & \quad C \\
\end{align*}
\]
Product-form stationary distribution

• Stationary distribution: *exactly* the same as OI queues!

\[ \pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}. \]
Product-form stationary distribution

- Stationary distribution: *exactly* the same as OI queues!

\[ \pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}. \]

- Proof: (partial) balance equations + normalization equation.
Product-form stationary distribution

• Stationary distribution: *exactly* the same as OI queues!

\[
\pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}.
\]

• Proof: (partial) balance equations + normalization equation.
• Hence, the P&S queue is a product-form queue.
Product-form stationary distribution

• Stationary distribution: *exactly* the same as OI queues!

\[
\pi(c_1, c_2, \ldots, c_n) = \pi(\emptyset) \prod_{p=1}^{n} \frac{\lambda_{c_p}}{\mu(c_1, \ldots, c_p)}.
\]

• Proof: (partial) balance equations + normalization equation.
• Hence, the P&S queue is a product-form queue.
• We also prove a simple stability condition (also valid for OI queues).
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:

\[ \begin{align*}
&\text{B} & \text{C} & \text{C} & \text{A} \\
&\mu(\cdot) \\
\end{align*} \]

• B always last in the upper queue, first in the lower queue: placement order.

The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:

\[
\begin{array}{c}
\text{A} \\
B & C & C & A \\
\mu(\cdot) \\
\nu(\cdot)
\end{array}
\]

- B always last in the upper queue, first in the lower queue: placement order.

- The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:

```
A
B C C A  \(\mu(\cdot)\)
\(\nu(\cdot)\)
```

- Always last in the upper queue, first in the lower queue: placement order.

The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:

A

B C C A

µ(·)

ν(·)

A B C

The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:

![Diagram of closed network of P&S queues](image)
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:

A

| A | C | C |

\[ \mu(\cdot) \]

| B |

\[ \nu(\cdot) \]

A -- B -- C
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:

\[
\begin{align*}
T & \text{ always last in the upper queue, first in the lower queue: } \\
\text{placement order.}
\end{align*}
\]

The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

• Tandem network of two P&S queues with the same swapping graph:
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:

\[
\begin{aligned}
A & \xrightarrow{\mu(\cdot)} C & C \\
\xleftarrow{\nu(\cdot)} B & & \end{aligned}
\]

A always last in the upper queue, first in the lower queue: placement order.

The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:

\[
\begin{array}{c}
A & C & C \\
\mu(\cdot) & & \\
\nu(\cdot) & B & \\
\end{array}
\]

- B always last in the upper queue, first in the lower queue: placement order.

The stationary distribution again has a product form (on a restricted space)!
Closed network of P&S queues

- Tandem network of two P&S queues with the same swapping graph:

  - B always last in the upper queue, first in the lower queue: *placement* order.
  - The stationary distribution again has a product form (on a restricted space)!

\[
\begin{align*}
\begin{array}{c}
\text{A} \quad \text{C} \quad \text{C} \\
\end{array}
& \quad \mu(\cdot) \\
\begin{array}{c}
\nu(\cdot) \\
\text{B} \\
\end{array}
\end{align*}
\]
From two queues to one queue

- The lower queue models “state-dependent arrivals” to the upper queue.
From two queues to one queue

- The lower queue models “state-dependent arrivals” to the upper queue.
- If the two queues are simple $\cdot/M/1$ queues, the upper queue can be seen as an $M/M/1$ queue with blocking.
From two queues to one queue

- The lower queue models "state-dependent arrivals" to the upper queue.
- If the two queues are simple \( \cdot /M/1 \) queues, the upper queue can be seen as an \( M/M/1 \) queue with blocking.
- This is very powerful:
  - Redundancy cancel-on-start and cancel-on-commit.
  - Hierarchical load-distribution algorithms.
Conclusion

Take away

• P&S queues broaden the family of product-form queues by allowing for intra-queue routing.
• Networks of P&S queues also have a product form.
• Paves the way for performance analysis of other algorithms.
Conclusion

Take away

• P&S queues broaden the family of product-form queues by allowing for intra-queue routing.
• Networks of P&S queues also have a product form.
• Paves the way for performance analysis of other algorithms.

Future works

• How big is the family of product-form queues?
• Are there other routing mechanisms that lead to a product form?
• Can we find other applications of P&S queues?