

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
KHOA CÔNG NGHỆ THÔNG TIN

PHÂN TÍCH HIỆU NĂNG QUY TRÌNH TỪ DỮ LIỆU BẰNG PHƯƠNG PHÁP XEM XÉT NHIỀU ĐỐI TƯỢNG

Sinh viên thực hiện:

Lê Thị Thùy Dương – MSSV: 20120063

Nguyễn Dương Tuấn Phương – MSSV: 20120166

Giảng viên hướng dẫn: ThS. Trần Trung Kiên và TS. Nguyễn Ngọc Thảo

THÀNH PHỐ HỒ CHÍ MINH, 31/07/2024

Mục lục

01

Giới thiệu đề tài

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp được đề xuất.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

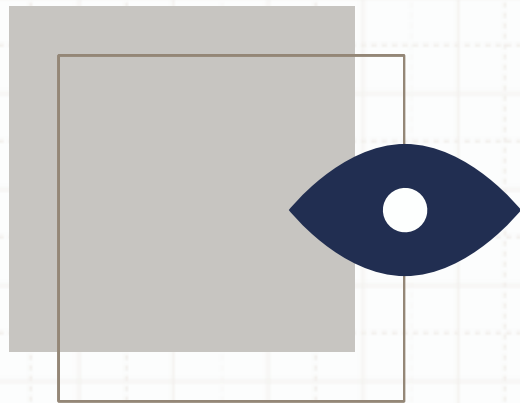
Tổng kết và hướng phát triển



01

GIỚI THIỆU ĐỀ TÀI

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp được đề xuất.





Khai thác quy trình

- Là một kỹ thuật được thiết kế để áp dụng các phương pháp chuyên biệt nhằm trích xuất các kiến thức hữu ích từ dữ liệu event log của hệ thống thông tin.
- Dữ liệu event log bao gồm danh sách các sự kiện được ghi lại theo thứ tự thời gian.





Dữ liệu đầu vào và yêu cầu của bài toán

Dữ liệu đầu vào

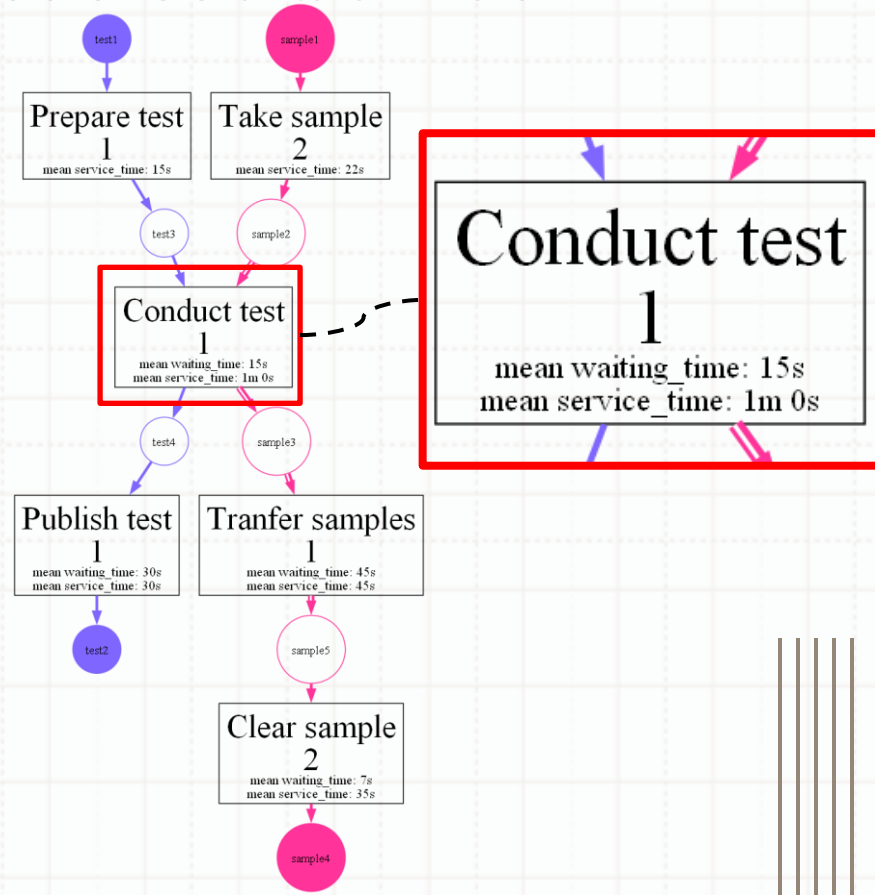
Dữ liệu event log của quy trình cần phân tích hiệu năng.

eID	Activity	Test	Sample	Start	End
e1	Prepare test	T1		0	15
e2	Take sample		S1	120	150
e3	Take sample		S2	150	165
e4	Conduct test	T1	S1, S2	180	240
e5	Publish test	T1		270	300
e6	Transfer samples		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400

Dữ liệu đầu vào và yêu cầu của bài toán

Yêu cầu bài toán

- Mô hình hóa được quy trình từ dữ liệu đầu vào.
- Tính toán các chỉ số hiệu năng cho từng bước trong quy trình.
- Hiển thị các chỉ số này cùng với mô hình quy trình tương ứng.





Khó khăn – ý nghĩa

Khó khăn và thách thức

- Yêu cầu dữ liệu đầy đủ và chất lượng cao.
- Các phương pháp tiếp cận truyền thống thường không thể hiện chính xác quy trình thực tế.

Ý nghĩa

- Trực quan hóa được quy trình làm việc từ dữ liệu đầu vào.
- Hỗ trợ người quản lý hiểu rõ hơn về quy trình hoạt động nghiệp vụ.
- Phát hiện các điểm thiếu hiệu quả trong quy trình vận hành.





Phương pháp truyền thống

Phương pháp truyền thống giải quyết bài toán phân tích hiệu năng quy trình từ dữ liệu đầu vào bằng việc tạo ra một mô hình quy trình Petri Net và sau đó các chỉ số hiệu năng được tính toán và ghi lại.

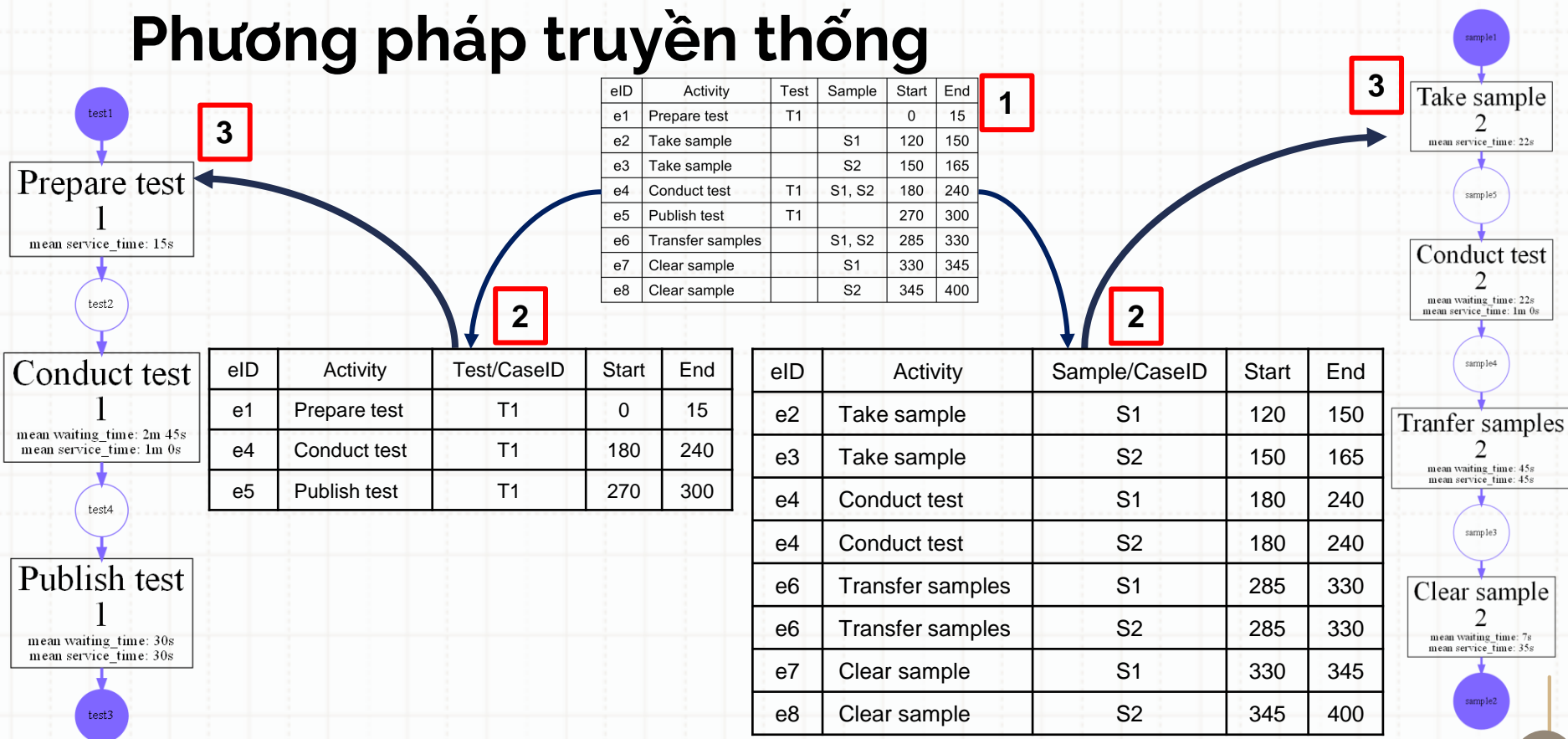
(P. T. G. Hornix, T.-I. A. J. M. M. Weijters, W.-D. P. M. E. D. Bra, W.-A. M. Voorhoeve, and P. H. Stramproy, "Performance Analysis of Business Processes through Process Mining". Master's thesis of P.T.G Hornix at Eindhoven University of Technology, 2007)



Phương pháp truyền thống

- Phương pháp này giả định rằng mỗi sự kiện trong dữ liệu event log chỉ thuộc về một trường hợp thực hiện quy trình (thường là chỉ liên quan đến một đối tượng).
- Với trường hợp một sự kiện liên quan đến nhiều đối tượng: cần phải tiền xử lý dữ liệu đầu vào để đúng với giả định.

Phương pháp truyền thống



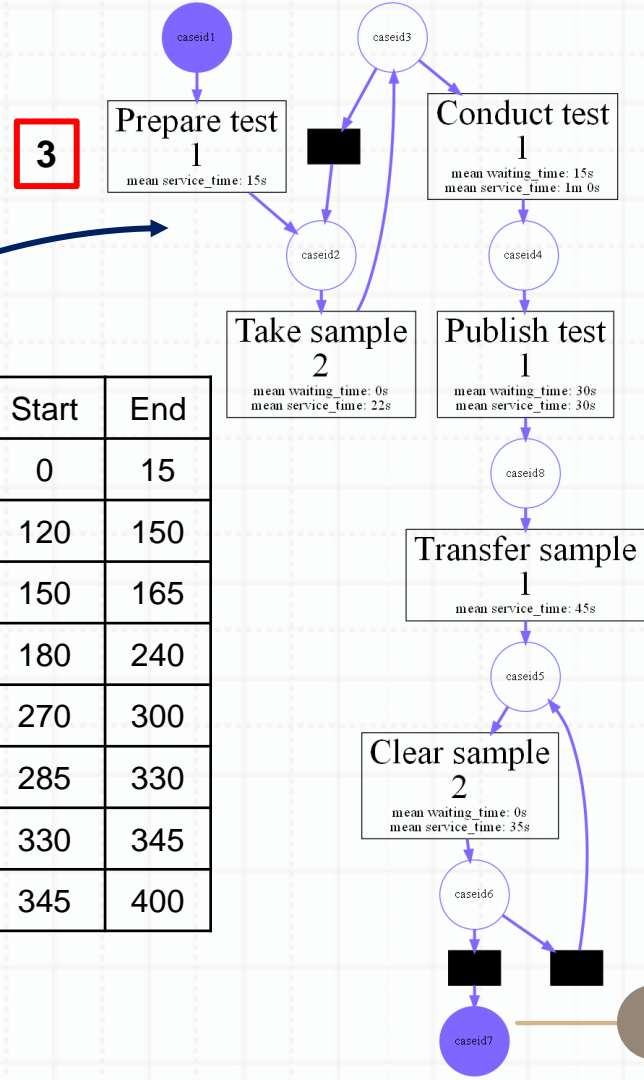
Phương pháp truyền thống

1

eID	Activity	Test	Sample	Start	End
e1	Prepare test	T1		0	15
e2	Take sample		S1	120	150
e3	Take sample		S2	150	165
e4	Conduct test	T1	S1, S2	180	240
e5	Publish test	T1		270	300
e6	Transfer samples		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400

2

eID	Activity	CaseID	Start	End
e1	Prepare test	Case1	0	15
e2	Take sample	Case1	120	150
e3	Take sample	Case1	150	165
e4	Conduct test	Case1	180	240
e5	Publish test	Case1	270	300
e6	Transfer samples	Case1	285	330
e7	Clear sample	Case1	330	345
e8	Clear sample	Case1	345	400





Phương pháp truyền thống

Mặc dù có thể giải quyết bài toán, tuy nhiên phương pháp này vẫn có những mặt hạn chế:

- Việc tiền xử lý dữ liệu làm mất mát thông tin có thể sẽ làm cho:
 - Mô hình quy trình không đúng với thực tế.
 - Các chỉ số hiệu năng không chính xác.
- Các chỉ số hiệu năng có liên quan nhiều đối tượng không được xem xét tới.

Phương pháp xem xét nhiều đối tượng

- Nhận thấy những hạn chế này, một phương pháp mới đã được đề xuất: xem xét nhiều đối tượng.
- Tương tự như phương pháp truyền thống nhưng thay mô hình Petri Net thành mô hình Object-Centric Petri Net.

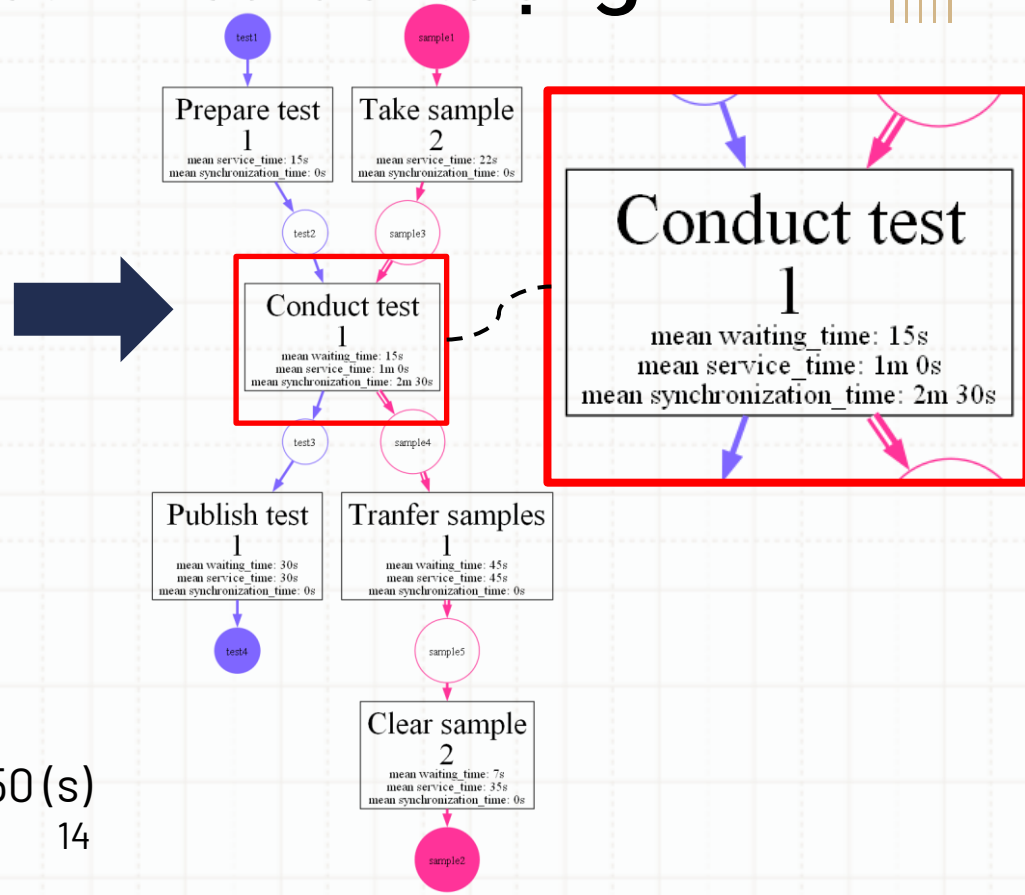
(G. Park, J. N. Adams, and W. M. P. van der Aalst, OPerA: "Object-Centric Performance Analysis". Springer International Publishing, 2022)

Phương pháp xem xét nhiều đối tượng

eID	Activity	Test	Sample	Start	End
e1	Prepare test	T1		0	15
e2	Take sample		S1	120	150
e3	Take sample		S2	150	165
e4	Conduct test	T1	S1, S2	180	240
e5	Publish test	T1		270	300
e6	Transfer samples		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400

Hoạt động Conduct test:

- Waiting time = $180 - 165 = 15$ (s)
- Service time = $240 - 180 = 60$ (s)
- Synchronization time = $165 - 15 = 150$ (s)



Phương pháp xem xét nhiều đối tượng

Dựa vào các thông tin đã trình bày, nhóm quyết định sẽ tập trung tìm hiểu sâu phương pháp xem xét nhiều đối tượng để giải quyết bài toán phân tích hiệu năng quy trình từ dữ liệu.

(G. Park, J. N. Adams, and W. M. P. van der Aalst, OPerA: "Object-Centric Performance Analysis". Springer International Publishing, 2022)

Mục lục

01

Giới thiệu đề tài

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp được đề xuất.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

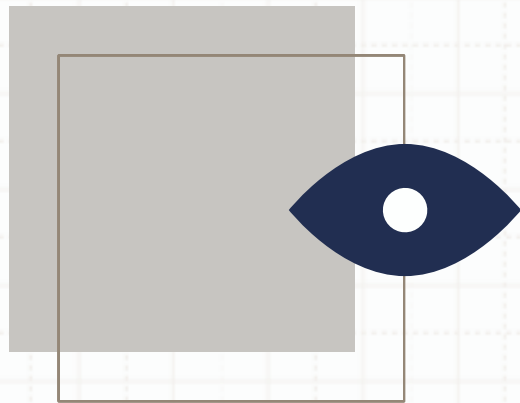
Tổng kết và hướng phát triển



02

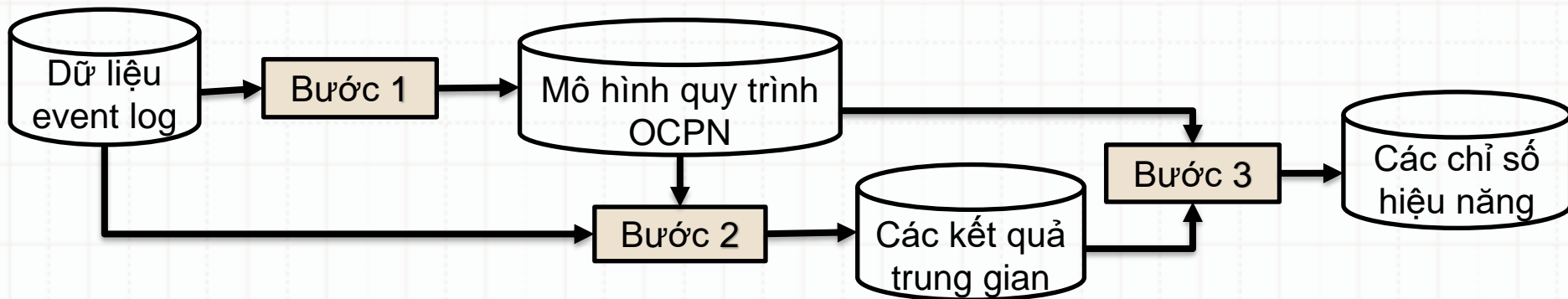
Phương pháp xem xét nhiều đối tượng

Trình bày chi tiết phương pháp xem xét nhiều đối tượng
Object-Centric Petri Net.



Xem xét nhiều đối tượng (Object-Centric)

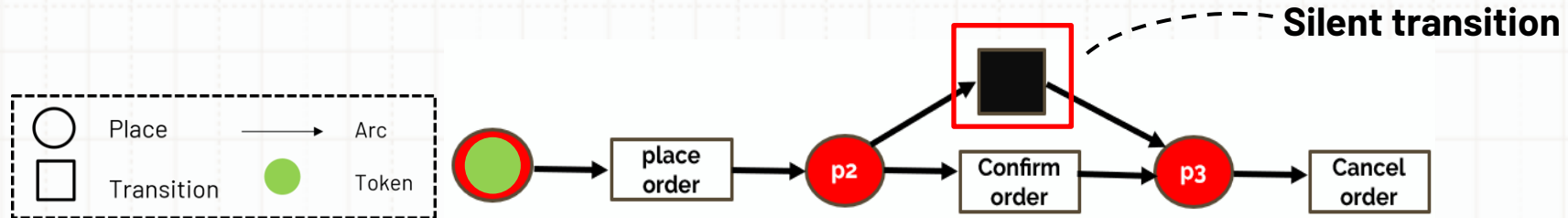
- **Bước 1:** tìm mô hình quy trình dạng Object-Centric Petri Net (OCPN) từ dữ liệu.
- **Bước 2:** tính các kết quả trung gian liên quan đến chỉ số hiệu năng.
- **Bước 3:** tính các chỉ số hiệu năng và biểu diễn trực quan lên mô hình quy trình.



Tìm mô hình quy trình có xem xét nhiều đối tượng từ dữ liệu

Petri Net là một loại mạng biểu diễn mô hình quy trình.

- Token sẽ di chuyển trong mạng, thể hiện cho quá trình thực hiện quy trình.
- Nút place đại diện cho một trạng thái trong quy trình.
- Nút transition đại diện cho quá trình chuyển tiếp token từ trạng thái này sang trạng thái khác.
- Cạnh kết nối giữa nút transition và nút place.



Tìm mô hình quy trình có xem xét nhiều đối tượng từ dữ liệu

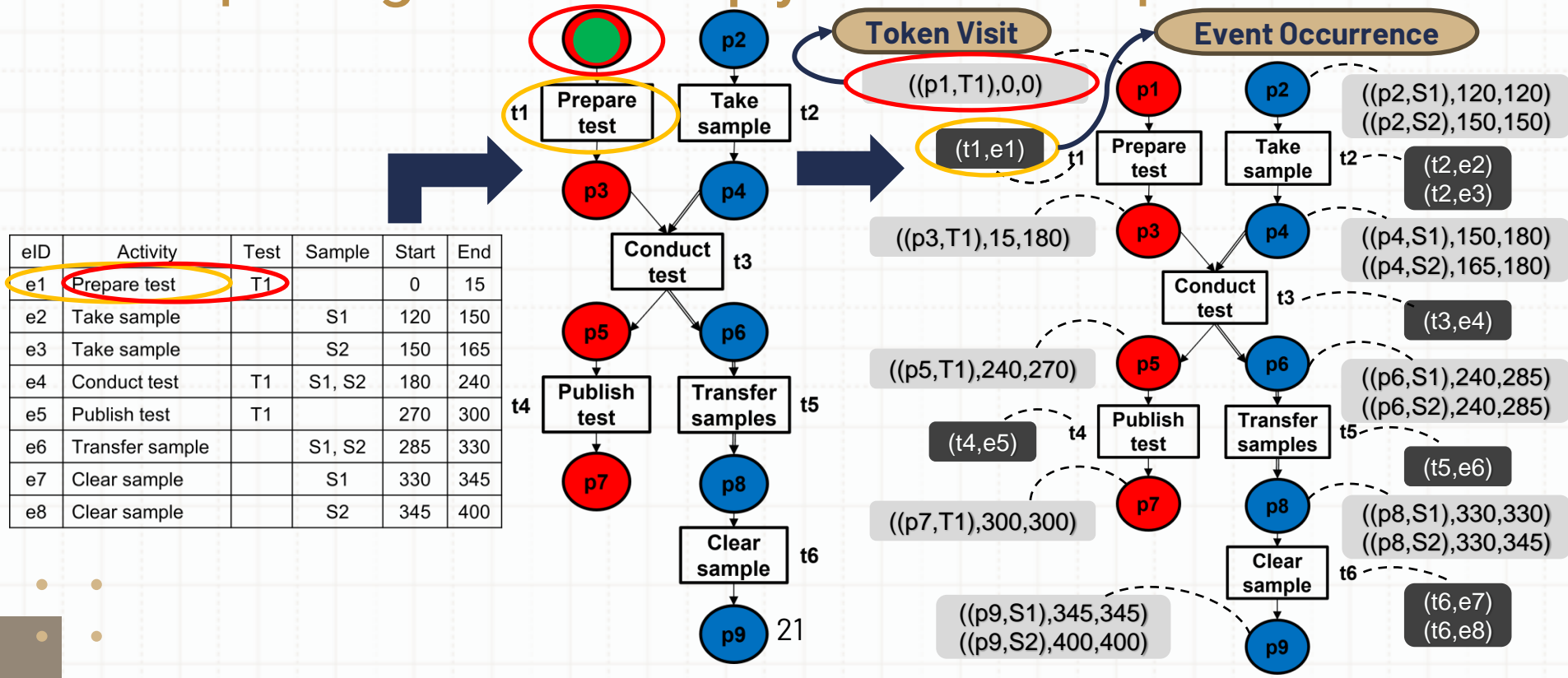
eID	Activity	Test	Sample	Start	End
e1	Prepare test	T1		0	15
e2	Take sample		S1	120	150
e3	Take sample		S2	150	165
e4	Conduct test	T1	S1, S2	180	240
e5	Publish test	T1		270	300
e6	Transfer samples		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400

Petri Net

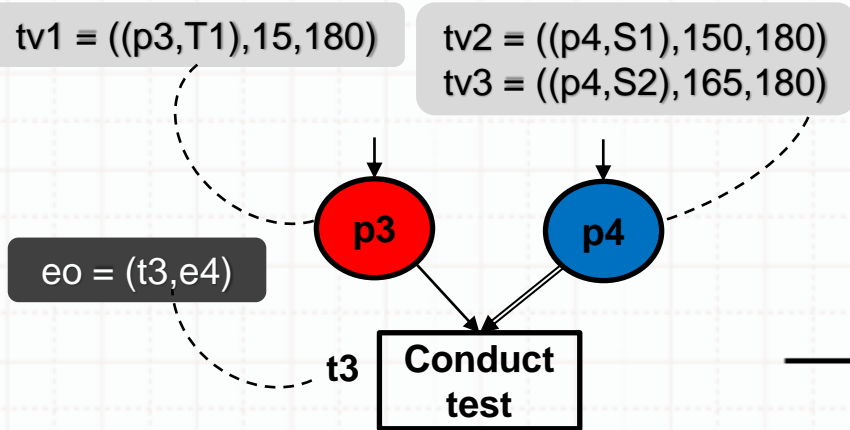
20

Object-Centric
Petri Net

Tính các kết quả trung gian liên quan đến chỉ số hiệu năng từ mô hình quy trình và dữ liệu

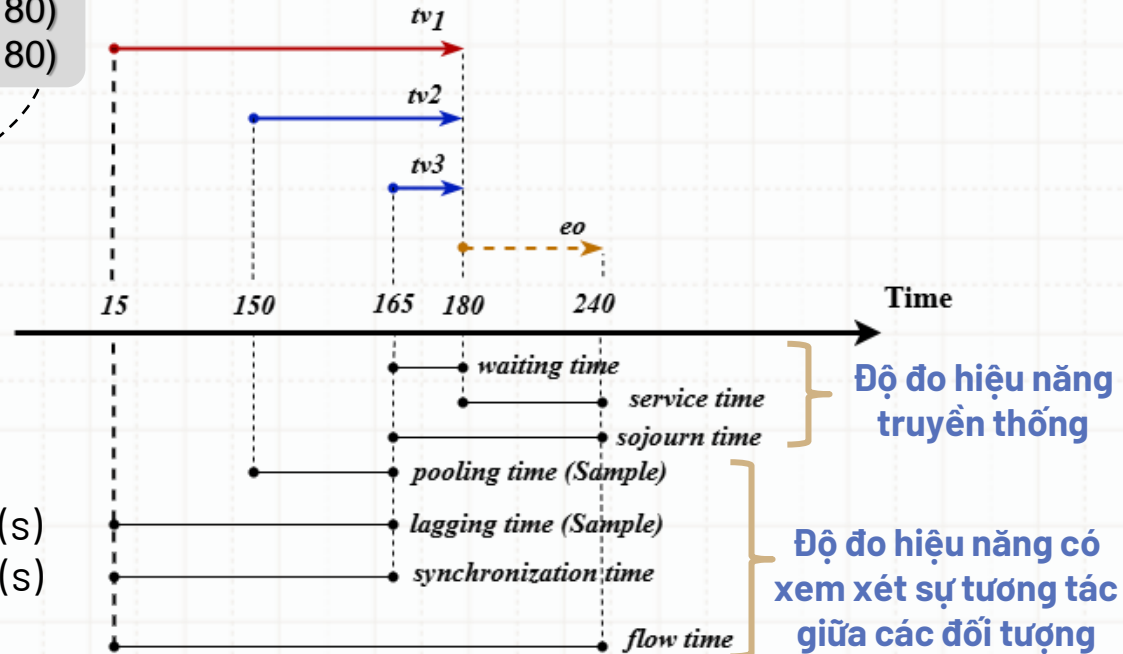


Tính các chỉ số hiệu năng từ các kết quả trung gian và mô hình quy trình



Ví dụ:

- Waiting time = $180 - 165 = 15$ (s)
- Pooling time (Sample) = $165 - 150 = 15$ (s)
- Lagging time (Sample) = $165 - 15 = 150$ (s)



Mục lục

01

Giới thiệu đề tài

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp được đề xuất.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

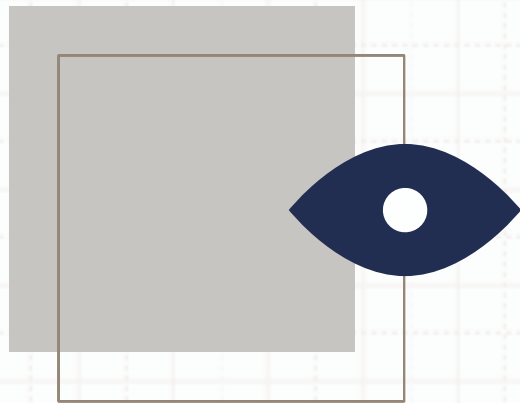
Tổng kết và hướng phát triển



03

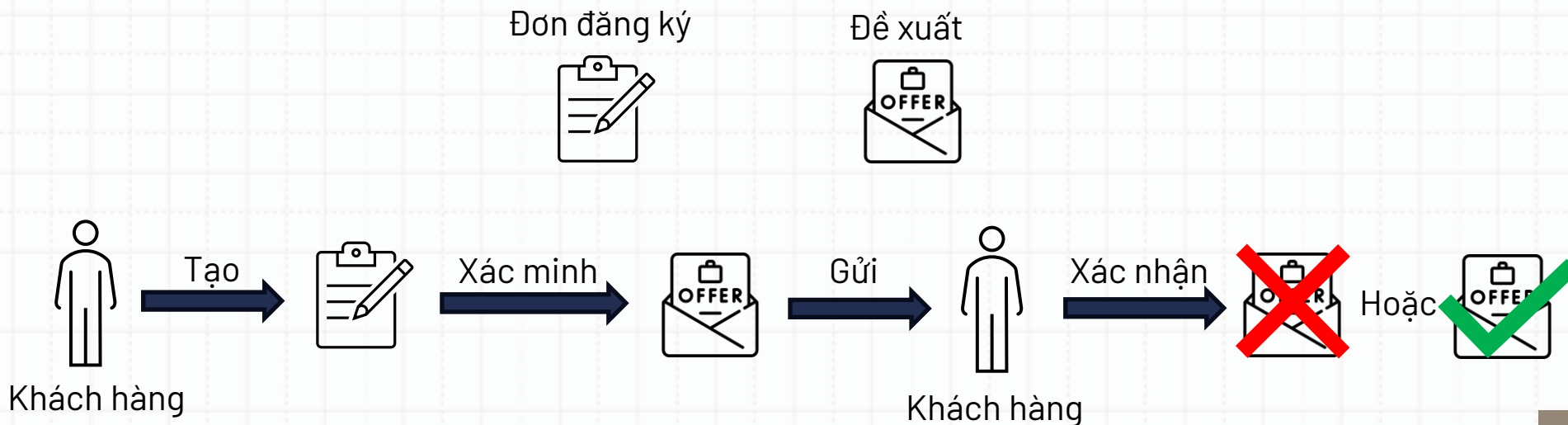
THỰC NGHIỆM

Kết quả thực nghiệm.



Bộ dữ liệu sử dụng

- Cung cấp bởi Hội thảo Quốc tế về Business Process Intelligence (BPI) lần thứ 13.
- Nội dung về quy trình đăng ký khoản vay của một viện Tài chính Hà Lan.





Rút trích dữ liệu

Với bộ dữ liệu BPI 2017:

- Tập trung vào các đề xuất cho vay bị hủy là chính.
- Chỉ giữ lại 10 hoạt động thường xuyên xảy ra nhất.
- Loại bỏ các hoạt động dư thừa.


Bộ dữ liệu BPI 2017 sau khi rút trích:

- Gồm 20,478 sự kiện.
- Hoàn thành bởi 1,682 đơn đăng ký và 3,573 đề xuất.



Các thí nghiệm

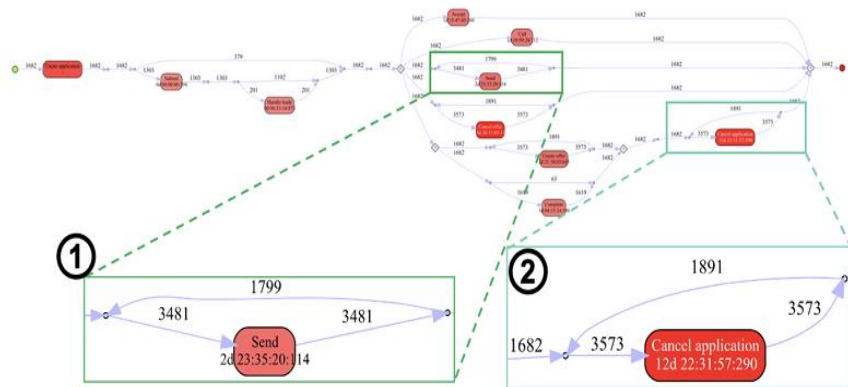
- Thí nghiệm 1: so sánh kết quả cài đặt khóa luận với bài báo gốc.
- Thí nghiệm 2: so sánh phương pháp xem xét nhiều đối tượng với phương pháp truyền thống (thí nghiệm mở rộng).
- Thí nghiệm 3: phân tích và tinh chỉnh mô hình quy trình phương pháp xem xét nhiều đối tượng (thí nghiệm mở rộng).



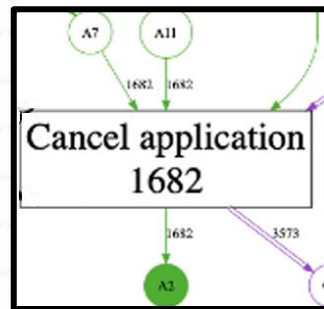
Thí nghiệm 1 - So sánh kết quả cài đặt của khóa luận với bài báo gốc

- Dùng phương pháp xem xét nhiều đối tượng để tiến hành thử nghiệm trên tập dữ liệu BPI 2017.
- Sử dụng kết quả có được để so sánh với kết quả trong bài báo gốc.

Kết quả bài báo gốc



a) Phương pháp truyền thống

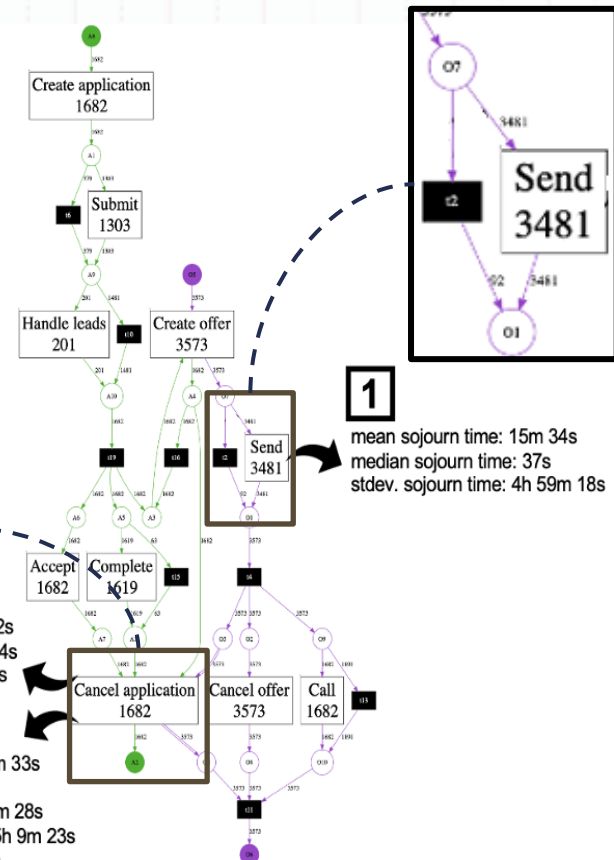


2

mean sojourn time: 31D 22h 11m 42s
median sojourn time: 30D 20h 2m 34s
stdev. sojourn time: 12D 1h 44m 15s

3

mean synchronization time: 4D 10h 45m 33s
median synchronization time: 2m 17s
stdev. synchronization time: 7D 21h 19m 28s
mean lagging time of Application: 3D 15h 9m 23s
mean lagging time of Offer: 19h 36m 9s
mean pooling time of Application: 4D 10h 41m 21s



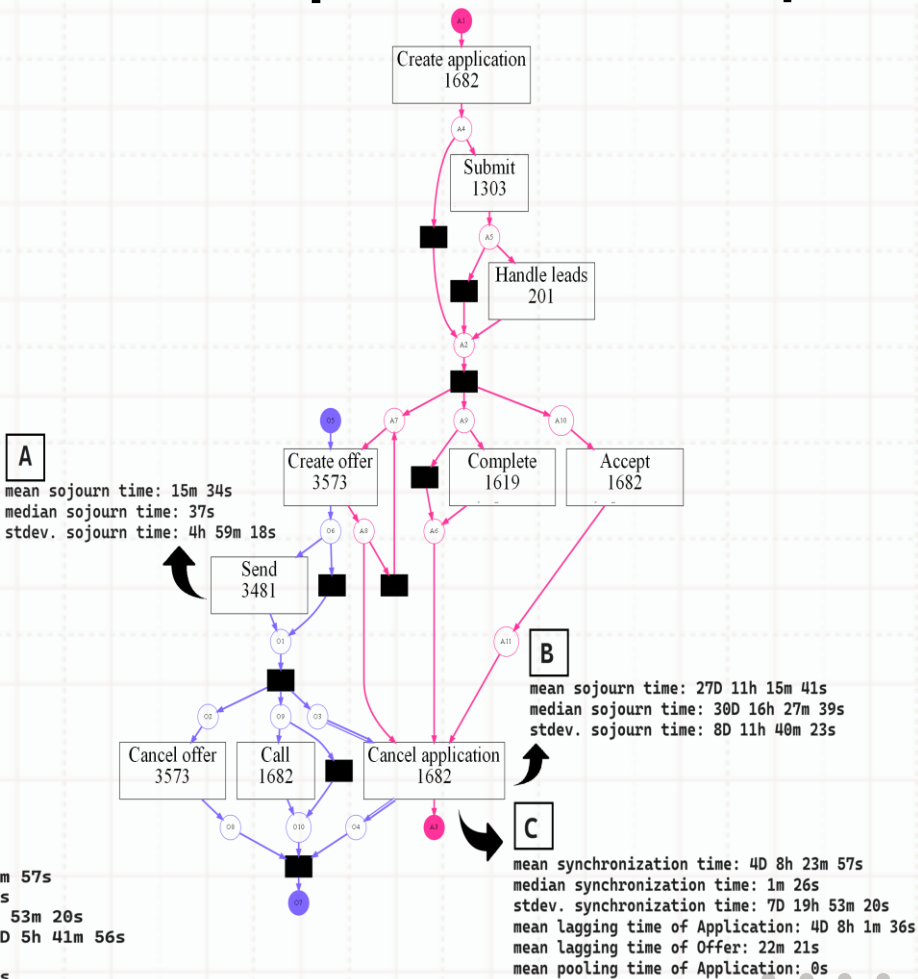
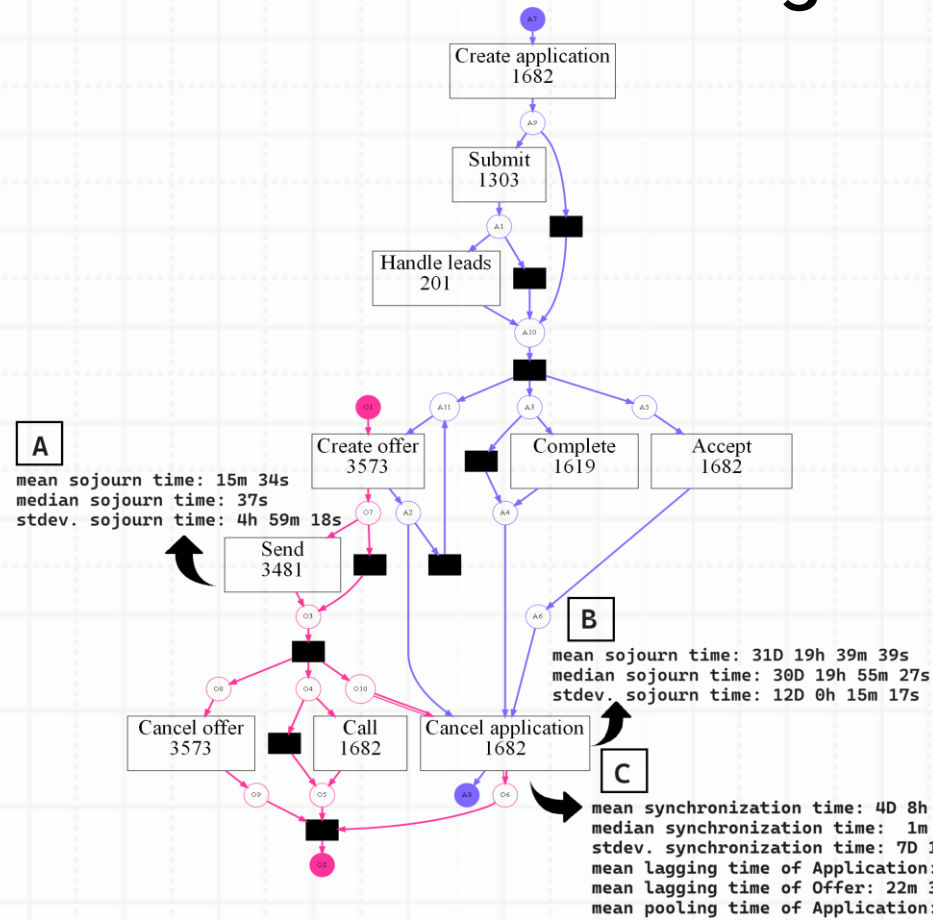
1

mean sojourn time: 15m 34s
median sojourn time: 37s
stdev. sojourn time: 4h 59m 18s

b) Phương pháp xem xét nhiều đối tượng

Kết quả phiên bản mã nguồn mới nhất của tác giả

Kết quả của khóa luận



Tổng hợp kết quả

Hoạt động	Độ đo	Kết quả bài báo	Phiên bản mã nguồn mới nhất của tác giả
Send	mean sojurn time	15m 34s	15m 34s
	median sojurn time	37s	37s
	stdev. sojurn time	4h 59m 18s	4h 59m 18s
Cancel Application	mean sojurn time	31D 22h 11m 42s	31D 19h 39m 39s ↓
	median sojurn time	30D 20h 2m 34s	30D 19h 55m 27s ↓
	stdev. sojurn time	12D 1h 44m 15s	12D 0h 15m 17s ↓
	mean synchronization time	4D 10h 45m 33s	4D 8h 23m 57s ↓
	median synchronization time	2m 17s	1m 26s ↓
	stdev. synchronization time	7D 21h 19m 28s	7D 19h 53m 20s ↓
	mean lagging time of Application	3D 15h 9m 23s	5D 5h 41m 56s ↑
	mean lagging time of Offer	19h 36m 9s	22m 31s ↓
	mean pooling time of Application	4D 10h 41m 21s	0s ↓

Tổng hợp kết quả

Hoạt động	Độ đo	Phiên bản mã nguồn mới nhất của tác giả	Cài đặt khóa luận
Send	mean sojurn time	15m 34s	15m 34s
	median sojurn time	37s	37s
	stdev. sojurn time	4h 59m 18s	4h 59m 18s
Cancel Application	mean sojurn time	31D 19h 39m 39s	27D 11h 15m 41s ↓
	median sojurn time	30D 19h 55m 27s	30D 16h 27m 39s ↓
	stdev. sojurn time	12D 0h 15m 17s	8D 11h 40m 23s ↓
	mean synchronization time	4D 8h 23m 57s	4D 8h 23m 57s
	median synchronization time	1m 26s	1m 26s
	stdev. synchronization time	7D 19h 53m 20s	7D 19h 53m 20s
	mean lagging time of Application	5D 5h 41m 56s	4D 8h 1m 36s ↓
	mean lagging time of Offer	22m 31s	22m 21s ↓
	mean pooling time of Application	0s	0s

Tổng hợp kết quả

Hoạt động	Độ đo	Phiên bản mã nguồn mới nhất của tác giả	Cài đặt khóa luận
Cancel Application	mean sojourn time	31D 19h 39m 39s	27D 11h 15m 41s ↓
	median sojourn time	30D 19h 55m 27s	30D 16h 27m 39s ↓
	stdev. sojourn time	12D 0h 15m 17s	8D 11h 40m 23s ↓

Thời gian lưu trú = thời điểm kết thúc sự kiện – thời điểm bắt đầu token visit muộn nhất


- Phiên bản mã nguồn mới nhất: token visit sớm nhất.
- Cài đặt khóa luận: token visit muộn nhất.

Tổng hợp kết quả

Hoạt động	Độ đo	Phiên bản mã nguồn mới nhất của tác giả	Cài đặt khóa luận
Cancel Application	mean lagging time of Application	5D 5h 41m 56s	4D 8h 1m 36s ↓
	mean lagging time of Offer	22m 31s	22m 21s ↓

Nếu thời gian trễ < 0 thì mặc định thời gian trễ = 0

- Phiên bản mã nguồn mới nhất: nếu thời gian trễ < 0 thì thời gian trễ = none (không tồn tại).
- Cài đặt khóa luận: nếu thời gian trễ < 0 thì thời gian trễ = 0.



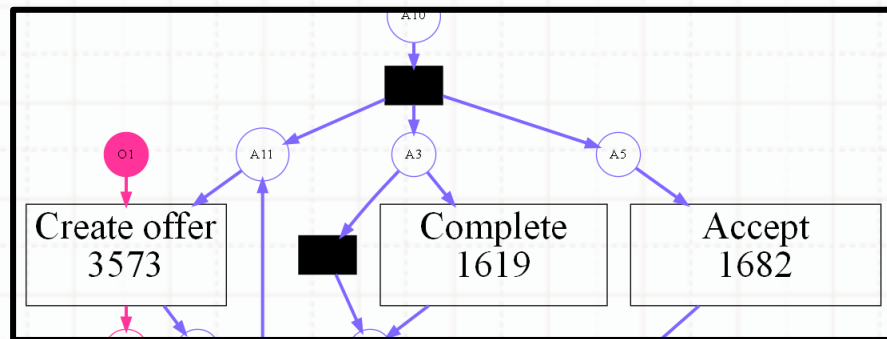
Thí nghiệm 3 - Phân tích và tinh chỉnh mô hình quy trình phương pháp xem xét nhiều đối tượng

(thí nghiệm mở rộng ngoài bài báo gốc)

Sau khi hoàn thành thí nghiệm 1, nhận thấy rằng mô hình quy trình hiện tại có một số điểm chưa hợp lý.

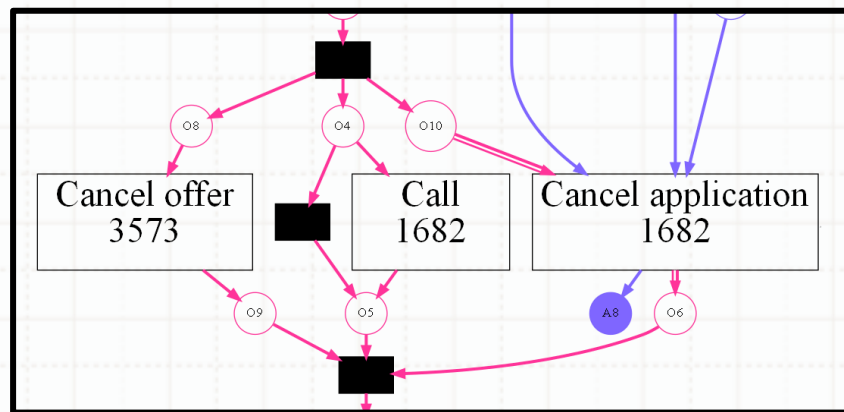
Điểm bất thường 1

- Đối với các hoạt động Accept, Complete và Create Offer: không có sự phân biệt thứ tự.
- Thứ tự xuất hiện phù hợp với bộ dữ liệu hiện tại nên là: hoạt động Accept diễn ra trước hoạt động Create Offer.



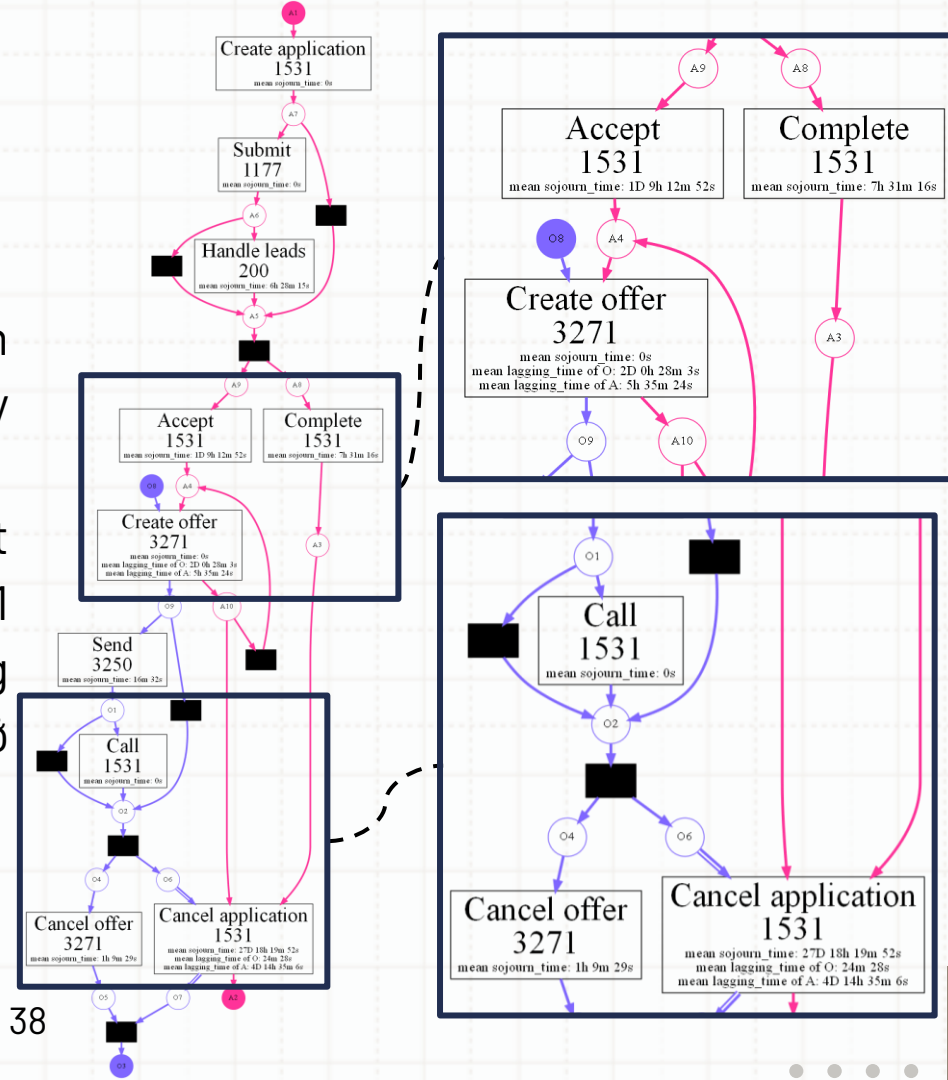
Điểm bất thường 2

- Các hoạt động Cancel Offer, Cancel Application và Call cũng không có sự phân biệt thứ tự.
- Thứ tự xuất hiện phù hợp với bộ dữ liệu hiện tại nên là: hoạt động Call diễn ra trước so với cả hai hoạt động Cancel Offer, Cancel Application.



Lọc dữ liệu – Kết quả

- Tiến hành lọc lại bộ dữ liệu hiện có và loại bỏ các trường hợp quy trình chưa phù hợp.
- Bộ dữ liệu sau khi lọc sẽ có tất cả là 1,513 đơn đăng ký và 3,271 đề xuất chiếm 91,02% trên tổng số 1,682 trường hợp quy trình ở bộ dữ liệu ban đầu.



Mục lục

01

Giới thiệu đề tài

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp được đề xuất.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

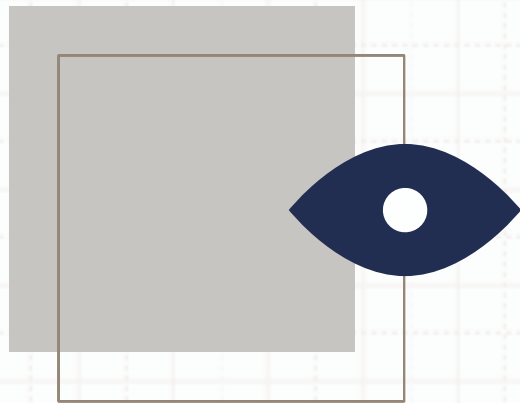
Tổng kết và hướng phát triển



04

TỔNG KẾT VÀ HƯỚNG PHÁT TRIỂN

Tổng kết những điểm chính và hướng phát triển trong tương lai.





Tổng kết

- Tìm hiểu và trình bày được phương pháp xem xét nhiều đối tượng.
- Cài đặt lại được từ đầu phương pháp và chỉnh sửa lại những điểm chưa chính xác ở cài đặt của tác giả.
- Có được các kết quả thí nghiệm nhằm so sánh và mở rộng.



Hướng phát triển

- Hỗ trợ phân tích hiệu năng ngay cả khi dữ liệu event log không tuân thủ các quy tắc nhất định.
- Phân tích hiệu năng tập trung vào đối tượng dựa trên dữ liệu event log, không phụ thuộc quá nhiều vào mô hình quy trình.

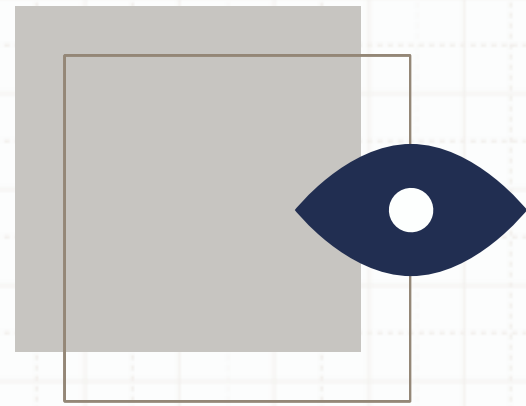



CẢM ƠN THẦY CÔ ĐÃ LẮNG NGHE



05

PHỤ LỤC





Thí nghiệm 2 - So sánh phương pháp xem xét nhiều đối tượng với phương pháp truyền thống

(thí nghiệm mở rộng ngoài bài báo gốc)

- So sánh phương pháp xem xét nhiều đối tượng với phương pháp truyền thống trên bộ dữ liệu BPI 2017.
- Sử dụng hai cách tiền xử lý dữ liệu khác nhau.

Cách tiền xử lý 1

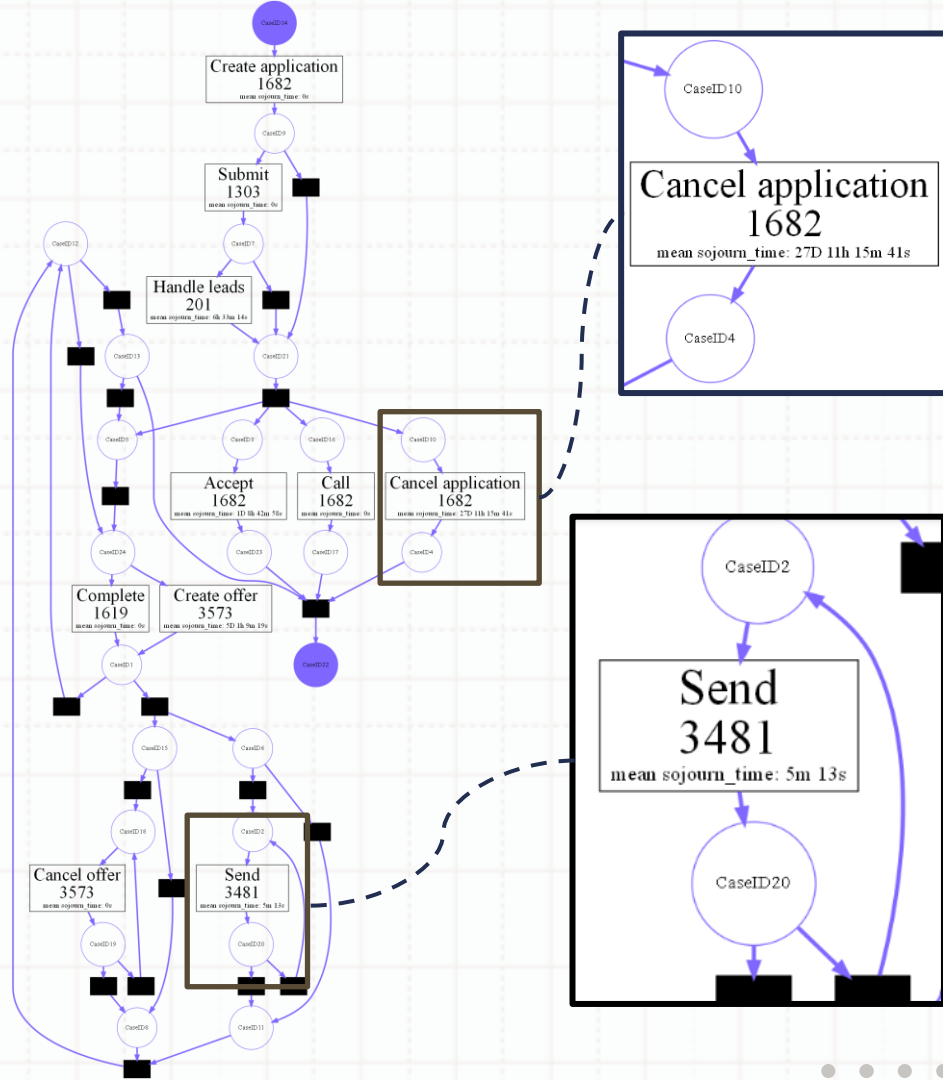
event_id	Application	event_activity	start_timestamp	timestamp	Offer
0	['A_1']	Create application	05:19.9	05:19.9	
1	['A_1']	Complete	22:55.9	45:11.2	
2	['A_1']	Accept	40:28.5	40:28.5	
3	['A_1']	Create offer	43:26.1	43:26.1	['Off_1']
4		Send	45:11.2	45:11.2	['Off_1']
5		Call	45:11.2	45:11.2	['Off_1']
6	['A_1']	Create offer	35:42.5	35:42.5	['Off_2']
7		Send	35:55.1	35:55.1	['Off_2']
8	['A_1']	Cancel application	00:48.8	00:48.8	['Off_1','Off_2']
9		Cancel offer	00:48.8	00:48.8	['Off_1']
10		Cancel offer	00:48.8	00:48.8	['Off_2']

Cách tiền xử lý 1

event_id	Application	event_activity	start_timestamp	timestamp	Offer	CaseID
0	['A_1']	Create application	05:19.9	05:19.9		Case_1
1	['A_1']	Complete	22:55.9	45:11.2		Case_1
2	['A_1']	Accept	40:28.5	40:28.5		Case_1
3	['A_1']	Create offer	43:26.1	43:26.1	['Off_1']	Case_1
4		Send	45:11.2	45:11.2	['Off_1']	Case_1
5		Call	45:11.2	45:11.2	['Off_1']	Case_1
6	['A_1']	Create offer	35:42.5	35:42.5	['Off_2']	Case_1
7		Send	35:55.1	35:55.1	['Off_2']	Case_1
8	['A_1']	Cancel application	00:48.8	00:48.8	['Off_1','Off_2']	Case_1
9		Cancel offer	00:48.8	00:48.8	['Off_1']	Case_1
10		Cancel offer	00:48.8	00:48.8	['Off_2']	Case_1

Cách tiền xử lý 1

- Hoạt động Cancel Application: không còn xuất hiện vòng lặp và thể hiện chính xác số lần xảy ra là 1,682 lần.
- Ở các hoạt động Send hay Cancel Offer: vẫn còn tồn tại vòng lặp.



Cách tiền xử lý 2

event_id	Application	event_activity	start_timestamp	timestamp	Offer
0	['A_1']	Create application	05:19.9	05:19.9	
1	['A_1']	Complete	22:55.9	45:11.2	
2	['A_1']	Accept	40:28.5	40:28.5	
3	['A_1']	Create offer	43:26.1	43:26.1	['Off_1']
4		Send	45:11.2	45:11.2	['Off_1']
5		Call	45:11.2	45:11.2	['Off_1']
6	['A_1']	Create offer	35:42.5	35:42.5	['Off_2']
7		Send	35:55.1	35:55.1	['Off_2']
8	['A_1']	Cancel application	00:48.8	00:48.8	['Off_1','Off_2']
9		Cancel offer	00:48.8	00:48.8	['Off_1']
10		Cancel offer	00:48.8	00:48.8	['Off_2']

Cách tiền xử lý 2

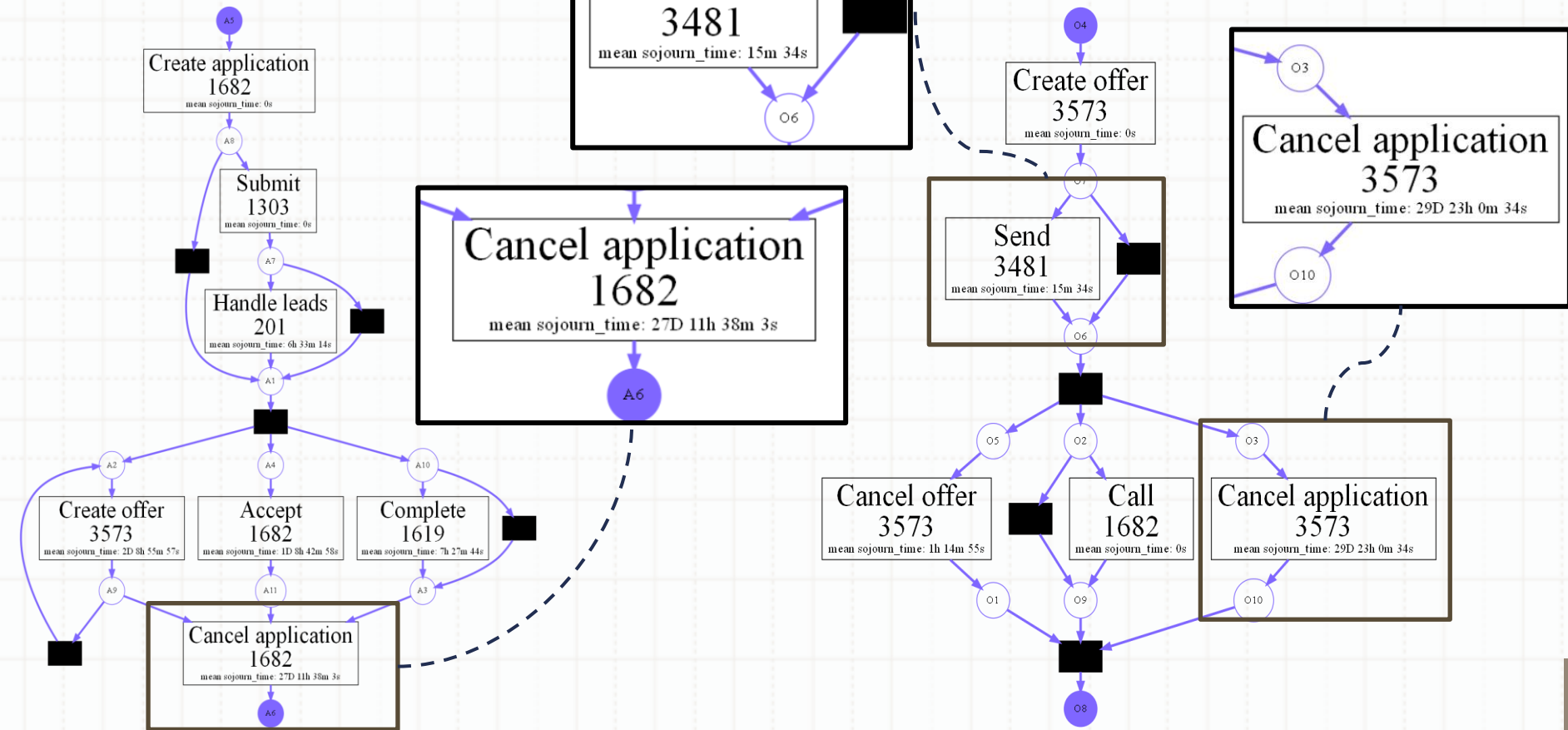
1

event_id	Application	event_activity	start_timestamp	timestamp
1	['A_1']	Create application	05:19.9	05:19.9
2	['A_1']	Complete	22:55.9	45:11.2
3	['A_1']	Accept	40:28.5	40:28.5
4	['A_1']	Create offer	43:26.1	43:26.1
5	['A_1']	Create offer	35:42.5	35:42.5
6	['A_1']	Cancel application	00:48.8	00:48.8

2

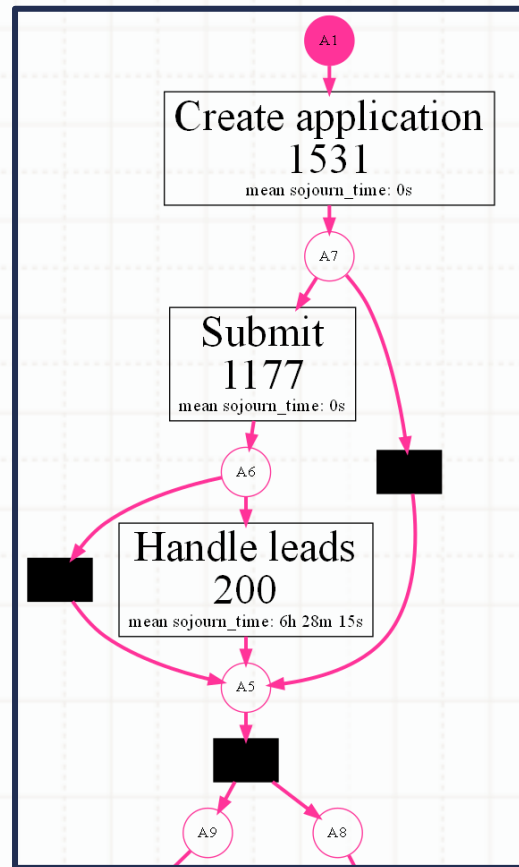
event_id	Offer	event_activity	start_timestamp	timestamp
1	['Off_1']	Create offer	43:26.1	43:26.1
2	['Off_1']	Send	45:11.2	45:11.2
3	['Off_1']	Call	45:11.2	45:11.2
4	['Off_2']	Create offer	35:42.5	35:42.5
5	['Off_2']	Send	35:55.1	35:55.1
6	['Off_1']	Cancel application	00:48.8	00:48.8
7	['Off_2']	Cancel application	00:48.8	00:48.8
8	['Off_1']	Cancel offer	00:48.8	00:48.8
9	['Off_2']	Cancel offer	00:48.8	00:48.8

Cách tiền xử lý 2



Ví dụ dữ liệu thực tế về Silent transition

- Dữ liệu có đầy đủ các hoạt động: Create application, Submit, Handle leads.
- Dữ liệu chỉ có hoạt động: Create application, Submit.
- Dữ liệu chỉ có hoạt động Create application.



event_id	Application	event_activity	start_timestamp	timestamp	Offer
292	['Application_1886382319']	Create application	51:28.5	51:28.5	[]
293	['Application_1886382319']	Complete	51:28.5	02:17.3	[]
294	['Application_1886382319']	Accept	53:10.5	53:10.5	[]
295	['Application_1886382319']	Create offer	58:49.4	58:49.4	['Offer_1658433857']
296	[]	Send	02:17.3	02:17.3	['Offer_1658433857']
297	[]	Call	02:17.4	02:17.4	['Offer_1658433857']
298	['Application_1886382319']	Create offer	11:11.0	11:11.0	['Offer_159153109']
299	[]	Send	11:22.3	11:22.3	['Offer_159153109']
300	['Application_1886382319']	Cancel application	03:05.3	03:05.3	['Offer_159153109', 'Offer_1658433857']
301	[]	Cancel offer	03:05.3	03:05.3	['Offer_159153109']
302	[]	Cancel offer	03:05.3	03:05.3	['Offer_1658433857']

