

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, NGÀY 31 THÁNG 07 NĂM 2024

Mục lục

01

Giới thiệu đề tài

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp giải quyết bài toán.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

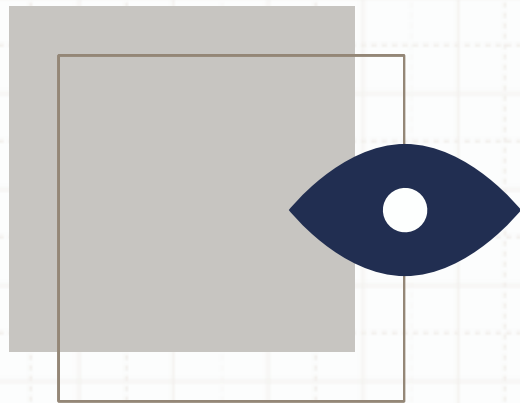
Tổng kết và hướng phát triển



01

GIỚI THIỆU ĐỀ TÀI

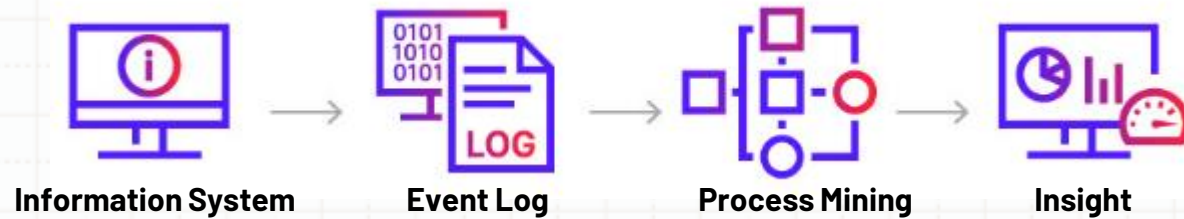
- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp giải quyết bài toán.





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 đầu vào và yêu cầu của bài toán

Dữ liệu đầu vào

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

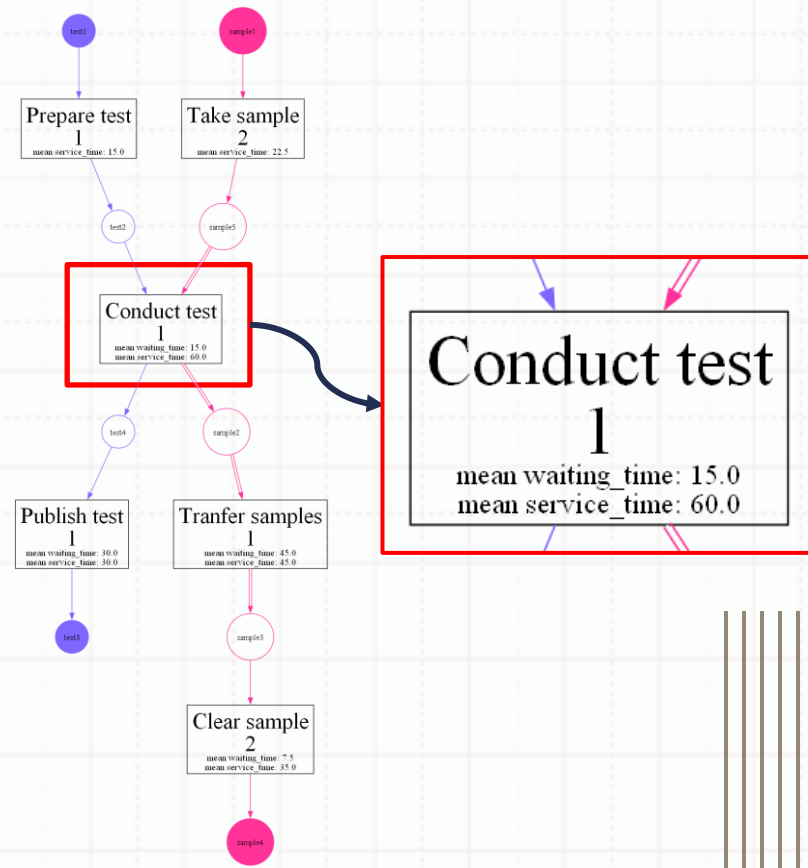
eID	Activity	Test	Sample	Start	End
e1	Prepare test	T1		0	15
e2	Take sample		S1	120	150
e3	Take sample		S2	150	165
e4	Conduct test	T1	S1, S2	180	165
e5	Publish test	T1		270	300
e6	Transfer sample		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400



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 (ví dụ như thời gian chờ trung bình, thời gian thực thi trung bình, ...) 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 về chất lượng của dữ liệu đầy đủ và chất lượng cao.
- Các phương pháp tiếp cận truyền thống để mô hình hóa quy trình thường không thể hiện đầy đủ những gì thực tế xảy ra bên trong của quy trình.

Ý nghĩa

- Hỗ trợ các doanh nghiệp, người quản lý hiểu rõ hơn về quy trình hoạt động nghiệp vụ.
- Phát hiện các nút place thiếu hiệu quả trong quy trình vận hành.
- Giúp tối ưu hóa, nâng cao chất lượng quy trình, tăng sự cạnh tranh và thích nghi nhanh chóng với sự biến đổi trong môi trường kinh doanh.



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.

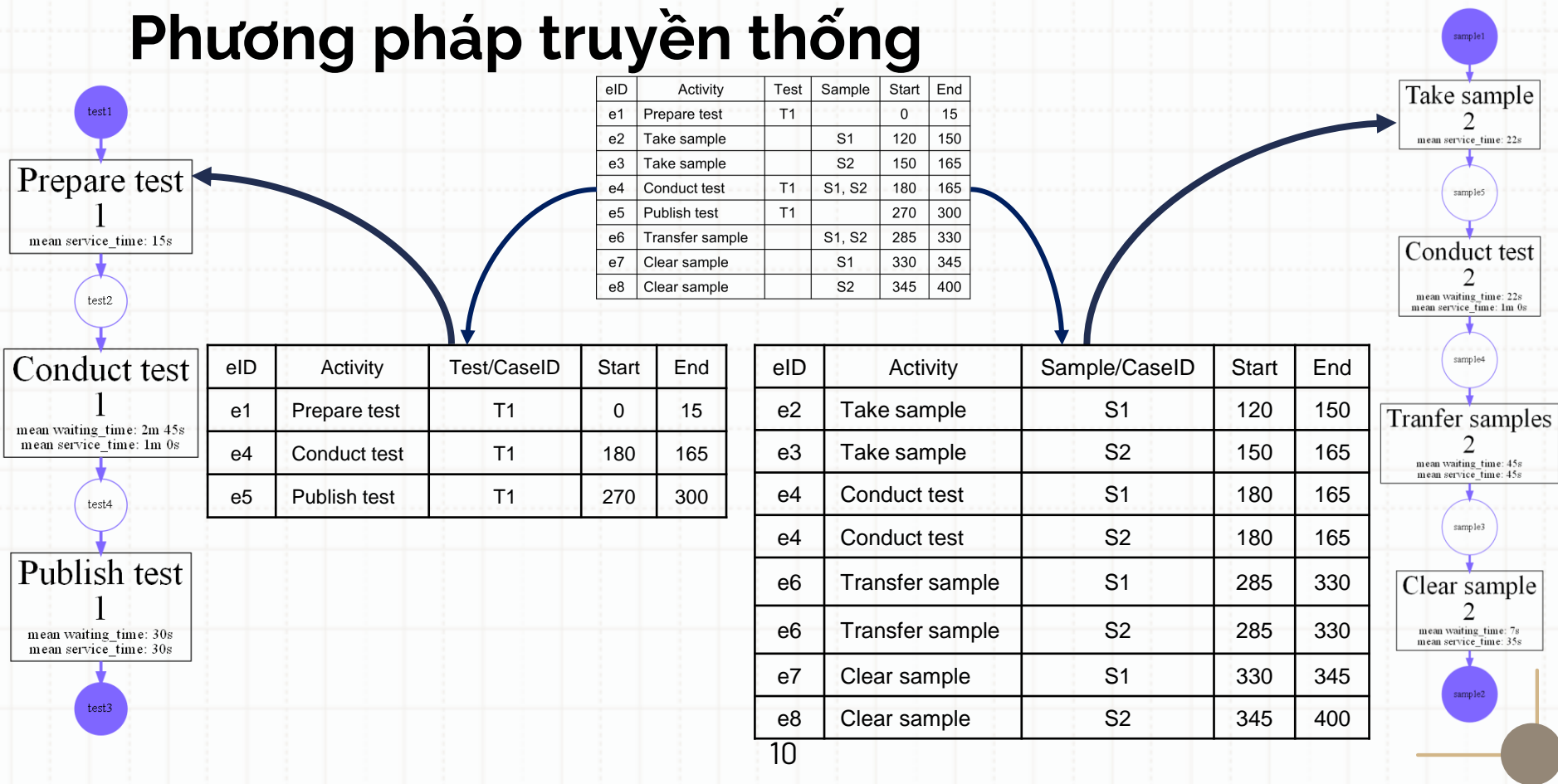
(P. T. G. Hornix, T.-I. A. J. M. M. Weijters, W.-D. P. M. E. D. Bra, W.-A. M. Voorhoeve, and P. H. Stramproy, "Performance Analysis of Business Processes through Process Mining". Master's thesis of P.T.G Hornix at Eindhoven University of Technology, 2007)



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.

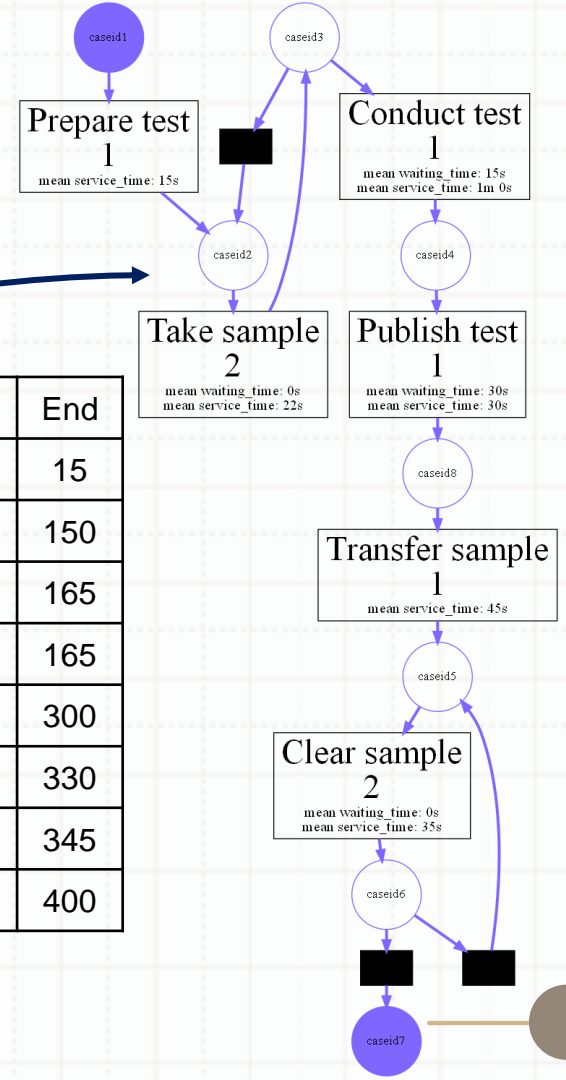
Phương pháp truyền thống



Phương pháp truyền thống

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	165
e5	Publish test	T1		270	300
e6	Transfer sample		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400

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	165
e5	Publish test	Case1	270	300
e6	Transfer sample	Case1	285	330
e7	Clear sample	Case1	330	345
e8	Clear sample	Case1	345	400

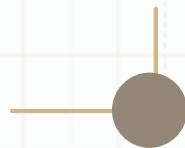




Phương pháp truyền thống

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 phản ánh được thực tế một cách chính xác và do đó có thể sẽ làm cho kết quả tính toán các chỉ số hiệu năng không được chính xác.
- Các chỉ số hiệu năng liên quan đến việc tương tác giữa các đối tượng không được xem xét tới.



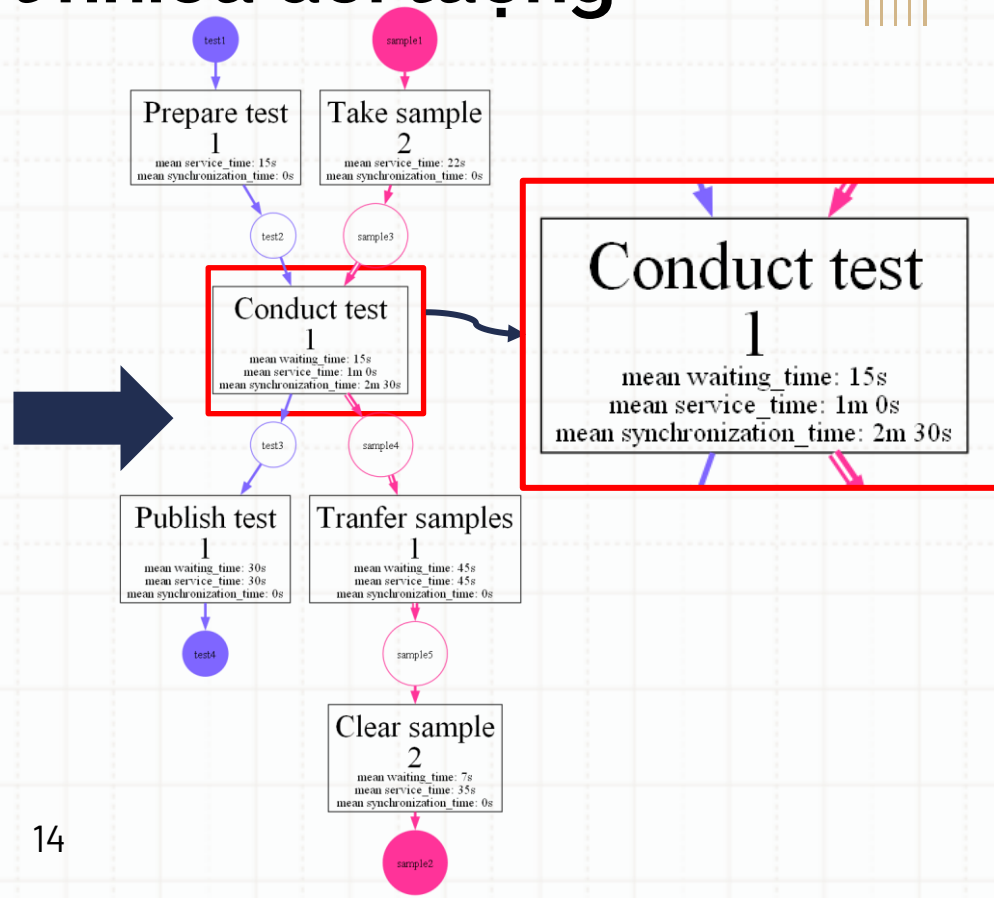
Phương pháp xem xét nhiều đối tượng

- Nhận thấy những hạn chế đó, thay vì sử dụng phương pháp truyền thống, một phương pháp tiếp cận mới đã được sử dụng: Object-Centric Petri Net.
- Sử dụng Object-Centric Petri Net như một phương pháp để mô hình quy trình dưới dạng một mạng lưới, qua đó phải thể hiện được sự tương tác giữa các đối tượng bằng đường mạng với những thông số hiệu năng trực quan hơn.

(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	165
e5	Publish test	T1		270	300
e6	Transfer sample		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400



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 giải quyết bài toán.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

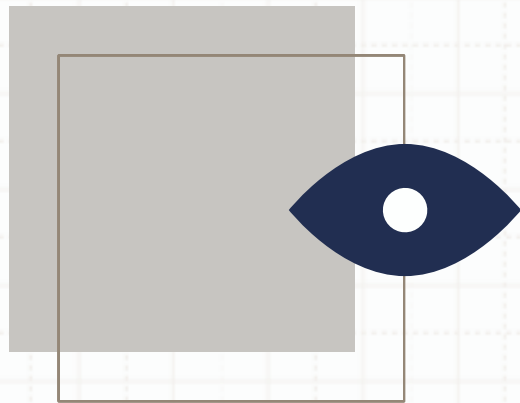
Tổng kết và hướng phát triển



02

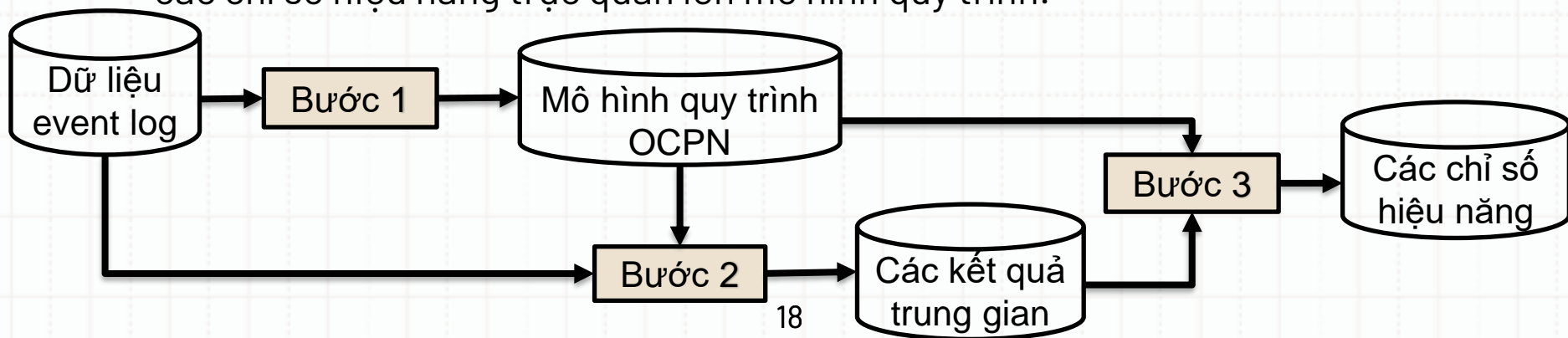
Phương pháp xem xét nhiều đối tượng

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



Xem xét nhiều đối tượng (Object-Centric)

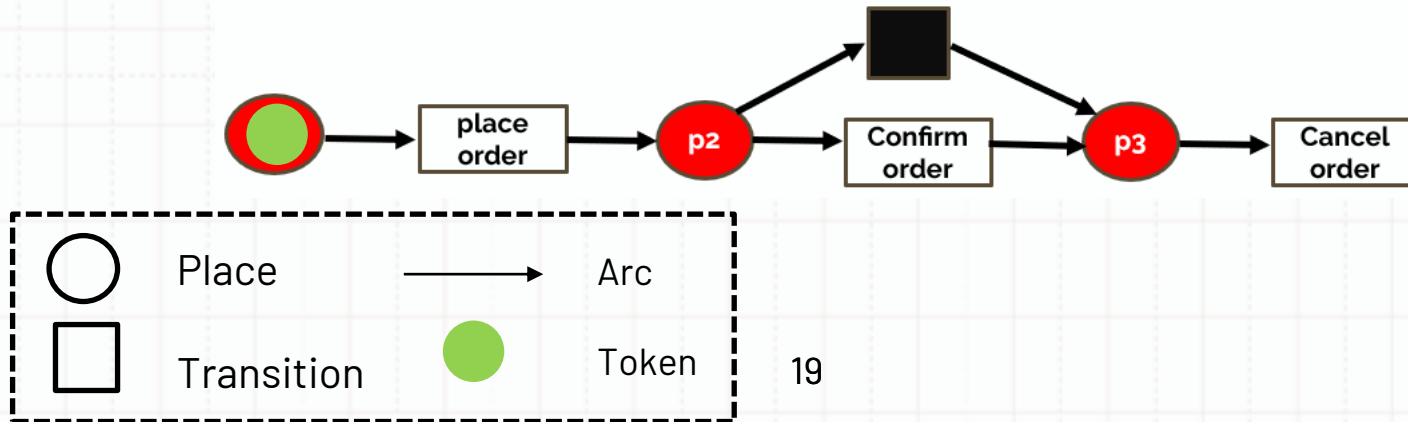
Phương pháp này bắt đầu bằng việc khám phá một mô hình quy trình dạng Object-Centric Petri Net (OCPN) từ dữ liệu event log. Tiếp theo, các sự kiện trong dữ liệu event log được phát lại trên mô hình quy trình này để cho ra các kết quả trung gian. Dựa vào đó, phương pháp sẽ đo lường, tính toán và ghi lại các chỉ số hiệu năng trực quan lên mô hình quy trình.



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

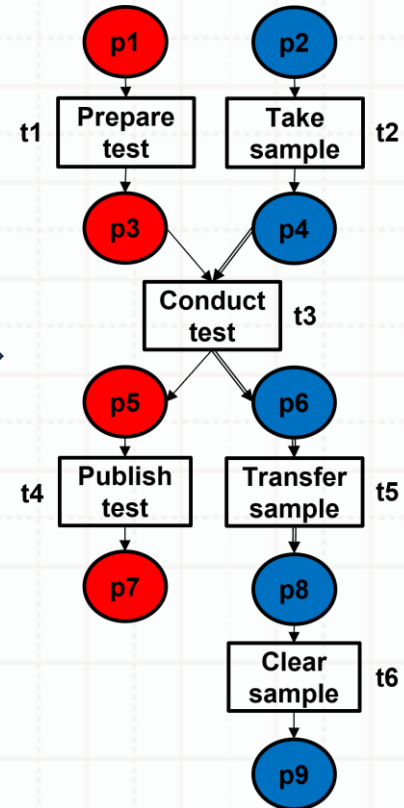
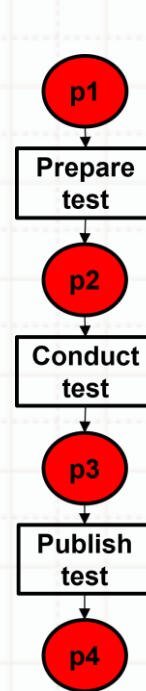
Petri Net là một mạng lưới tĩnh và không thay đổi.

- Các nút place đại diện cho một trạng thái trong quy trình, có thể chứa các token.
- Nút transition đại diện cho quá trình chuyển tiếp token từ trạng thái này sang trạng thái khác.

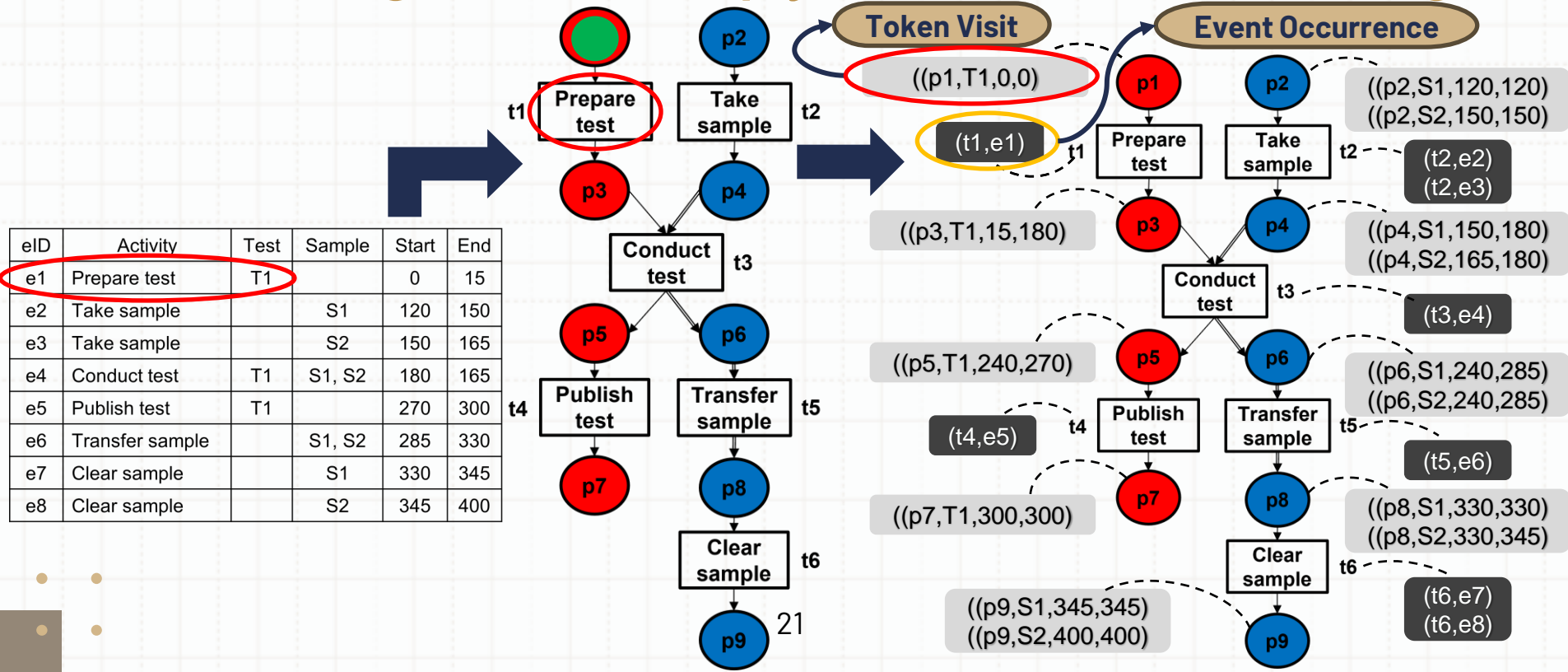


Tìm mô hình quy trình có xem xét nhiều đối tượng từ dữ liệu event log

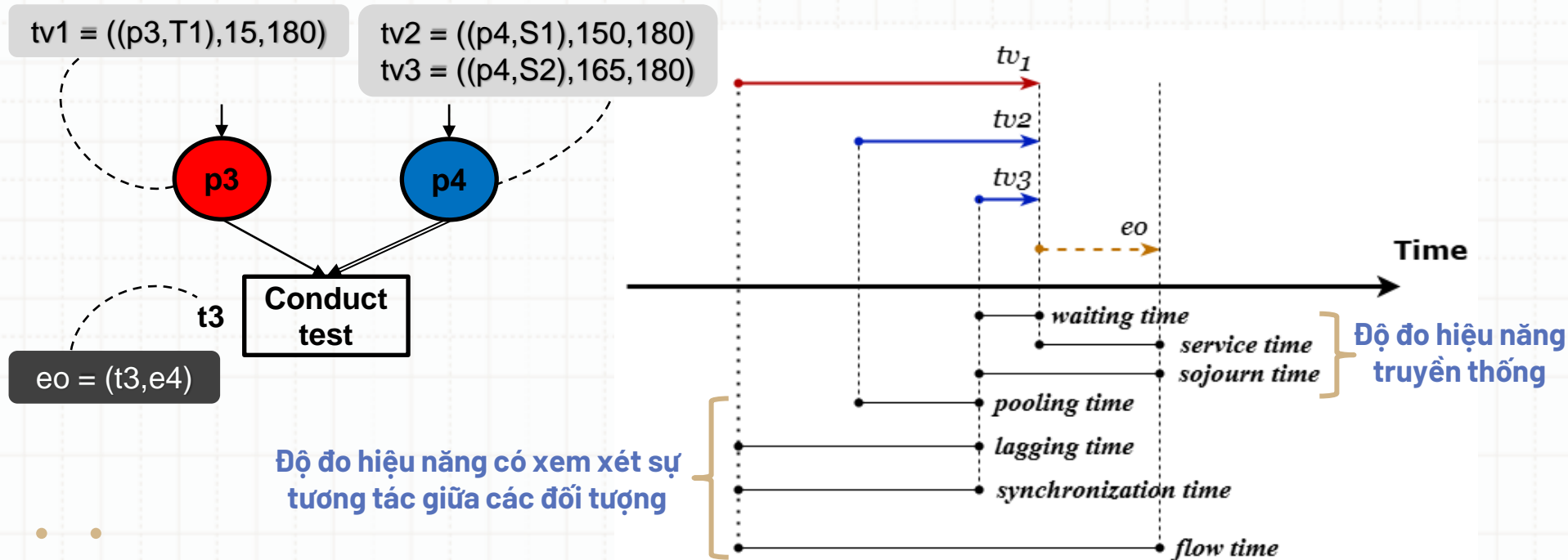
eID	Activity	Test	Sample	Start	End
e1	Prepare test	T1		0	15
e2	Take sample		S1	120	150
e3	Take sample		S2	150	165
e4	Conduct test	T1	S1, S2	180	165
e5	Publish test	T1		270	300
e6	Transfer sample		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400



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



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



Mục lục

01

Giới thiệu đề tài

- Bài toán phân tích hiệu năng quy trình.
- Các phương pháp giải quyết bài toán.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

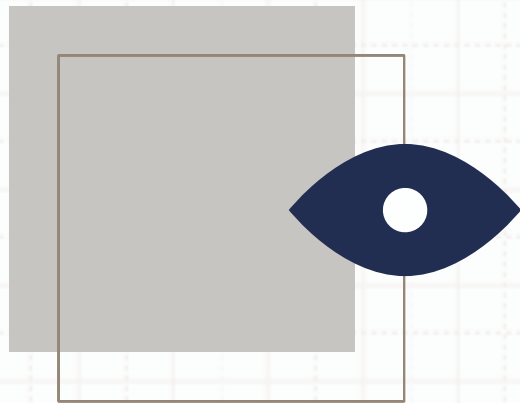
Tổng kết và hướng phát triển



03

THỰC NGHIỆM

Kết quả thực nghiệm.





Bộ dữ liệu sử dụng

Bộ dữ liệu về quy trình đăng ký khoản vay của một viện Tài chính Hà Lan, được cung cấp bởi Hội thảo Quốc tế về Business Process Intelligence (BPI) lần thứ 13 diễn ra vào năm 2017.

Bộ dữ liệu về quy trình đăng ký khoản vay của một viện Tài chính Hà Lan:

- Có 2 đối tượng chính tồn tại: đơn đăng ký (Application) và đề xuất (Offer).
- Một đơn đăng ký có thể có một hoặc nhiều đề xuất.
- Khách hàng tạo đơn đăng ký bằng cách đến trực tiếp ngân hàng hoặc sử dụng hệ thống trực tuyến.
- Sau khi hoàn thiện và xác minh hồ sơ, đề xuất khoản vay cho khách hàng sẽ được gửi và có thể xác nhận thông qua cuộc gọi, email. Một đề xuất có thể được chấp nhận hoặc hủy bỏ.




Rút trích dữ liệu

Với bộ dữ liệu BPI 2017:

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

Bộ dữ liệu BPI 2017 sau khi rút trích:

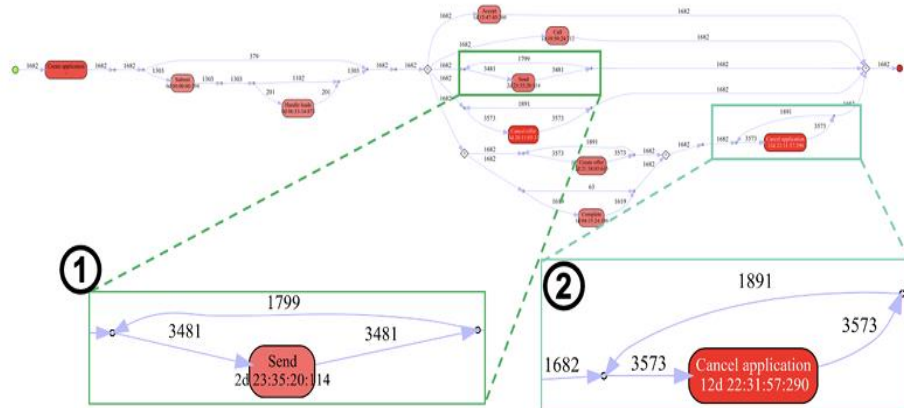
- Gồm 20,478 sự kiện.
- Hoàn thành bởi 1,682 đơn đăng ký và 3,573 đề xuất.



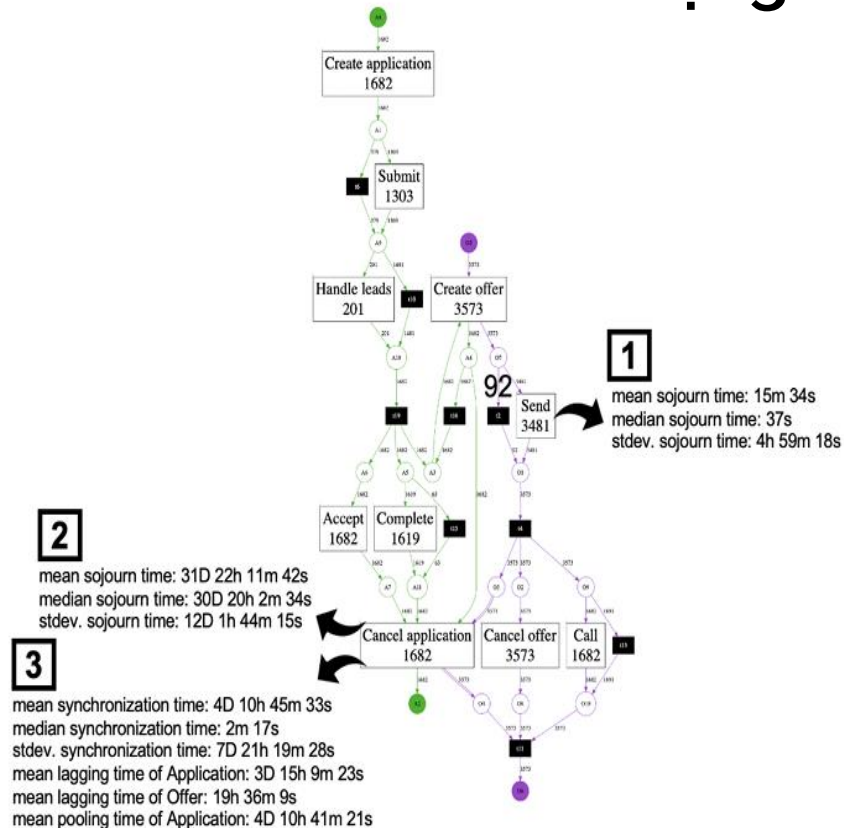
Thí nghiệm 1 - So sánh kết quả cài đặt của khóa luận với bài báo gốc

Ở thí nghiệm đầu tiên, chúng em sẽ sử dụng phương pháp xem xét nhiều đối tượng để tiến hành thử nghiệm trên tập dữ liệu “BPI 2017”. Sau đó chúng em sẽ sử dụng kết quả có được để so sánh với kết quả trong bài báo gốc.

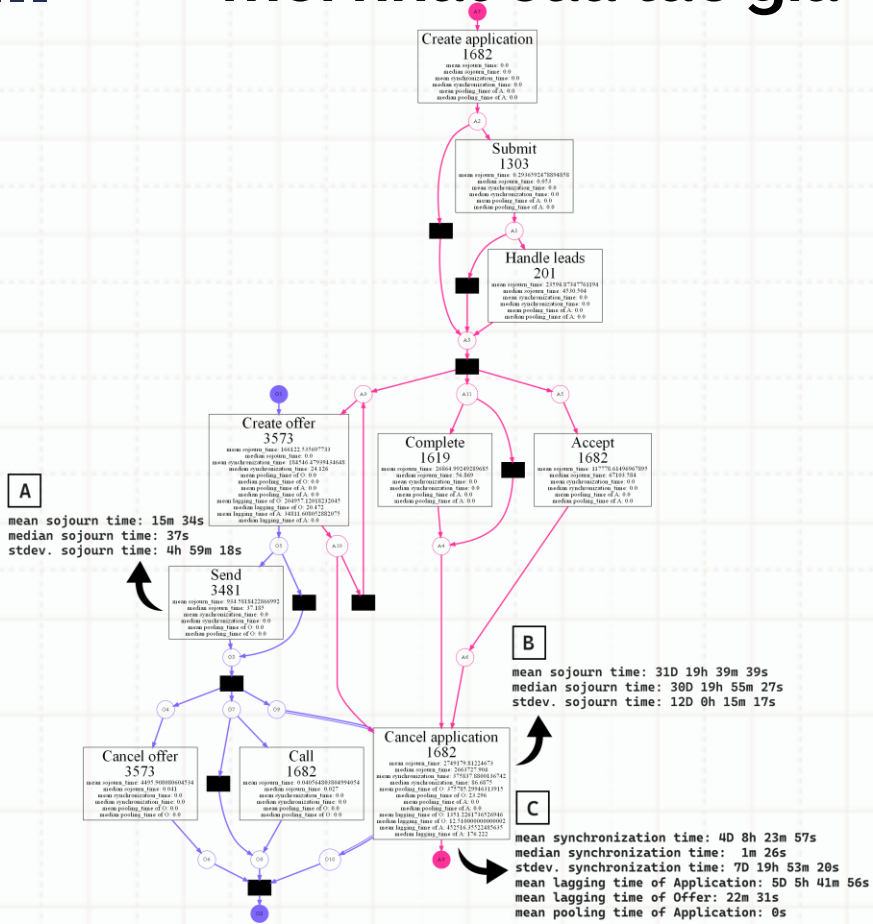
Kết quả bài báo gốc sử dụng phương pháp truyền thống



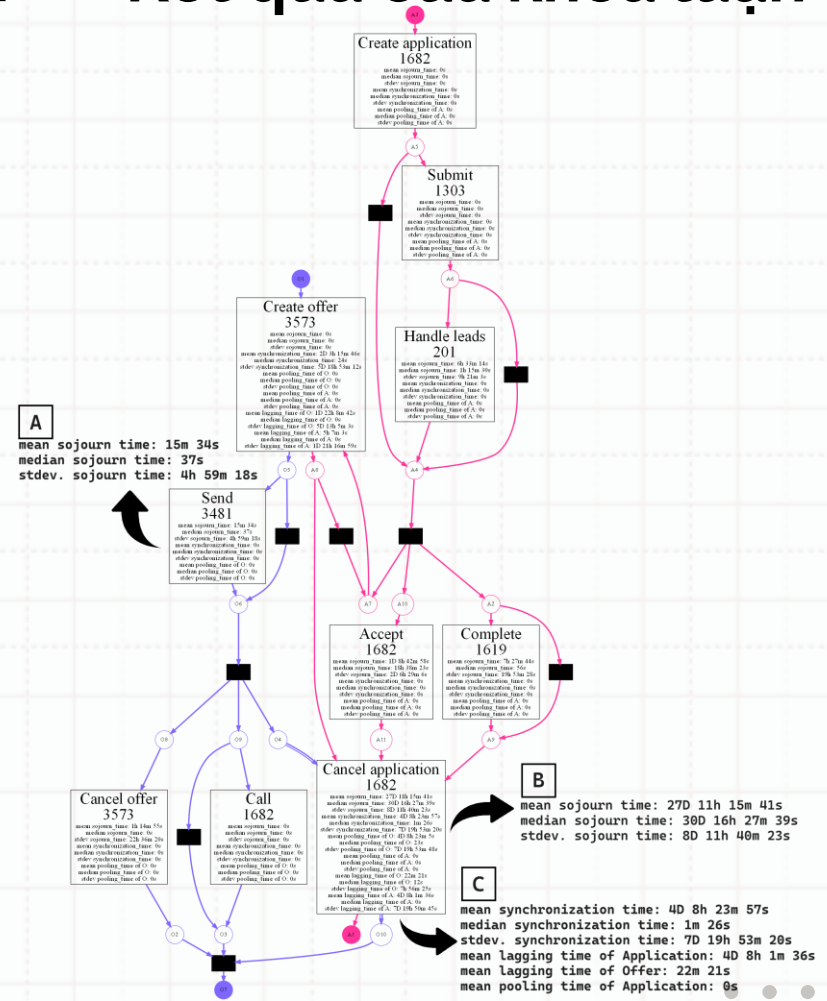
Kết quả bài báo gốc sử dụng phương pháp xem xét nhiều đối tượng



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



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

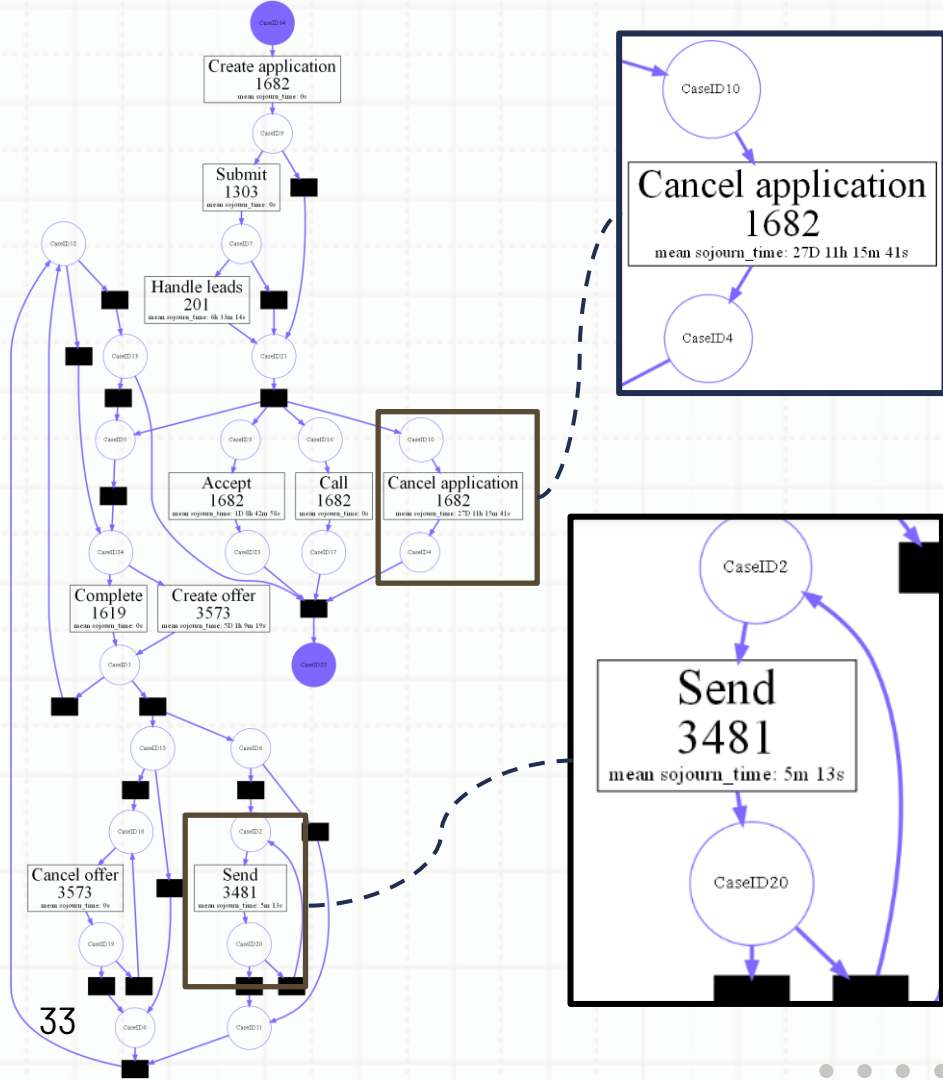
(thí nghiệm mở rộng ngoài bài báo gốc)

Trong thí nghiệm 2, chúng em sẽ tiến hành so sánh các kết quả từ phương pháp xem xét nhiều đối tượng với phương pháp truyền thống trên bộ dữ liệu BPI 2017 theo cài đặt của khóa luận. Sử dụng hai cách tiền xử lý dữ liệu khác nhau:

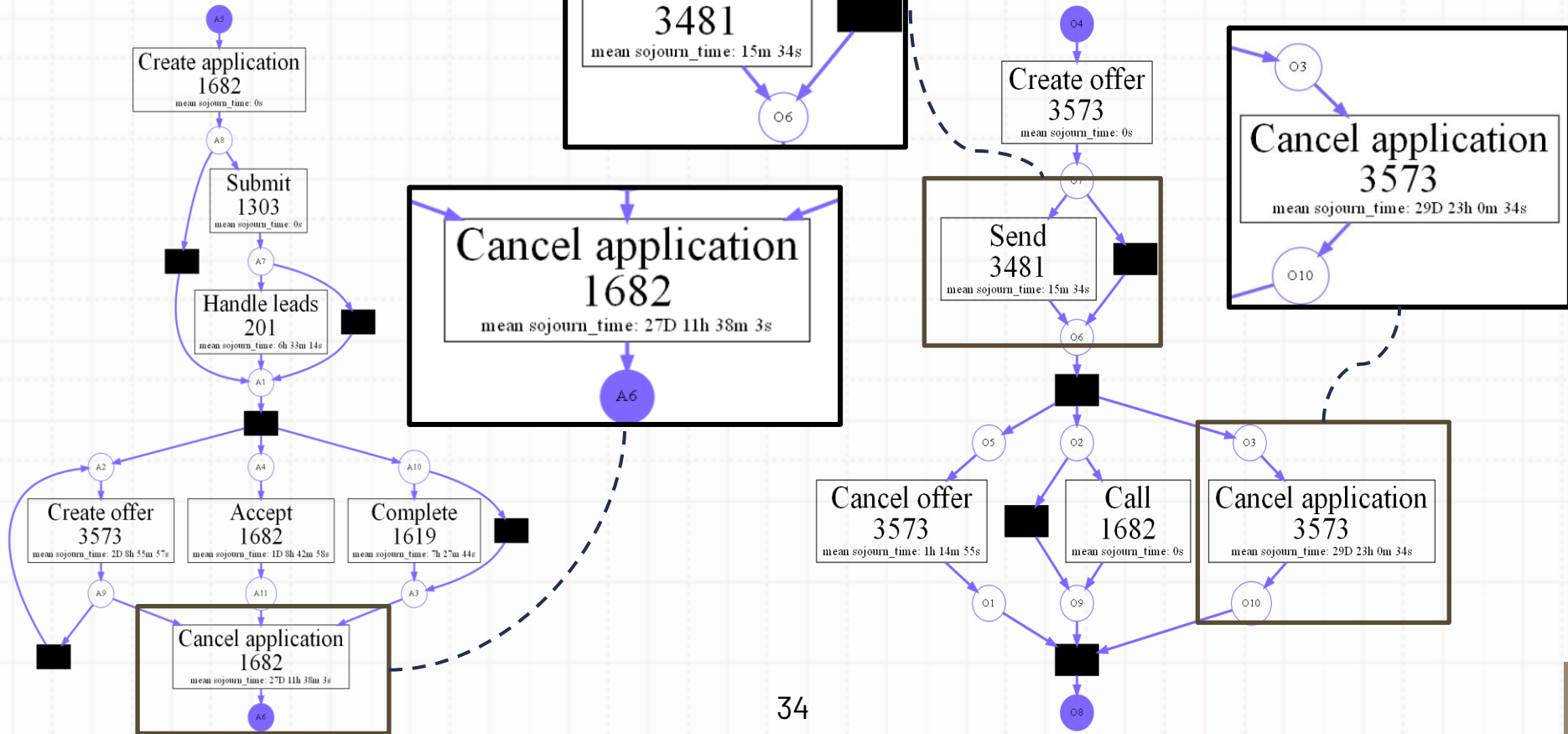
- Cách thứ nhất: phương pháp tiền xử lý dữ liệu event log sao cho mỗi trường hợp thực hiện quy trình sẽ bao gồm nhiều đối tượng.
- Cách thứ hai: phương pháp tiền xử lý dữ liệu event log sao cho mỗi trường hợp thực hiện quy trình chỉ bao gồm một đối tượng thực hiện.


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



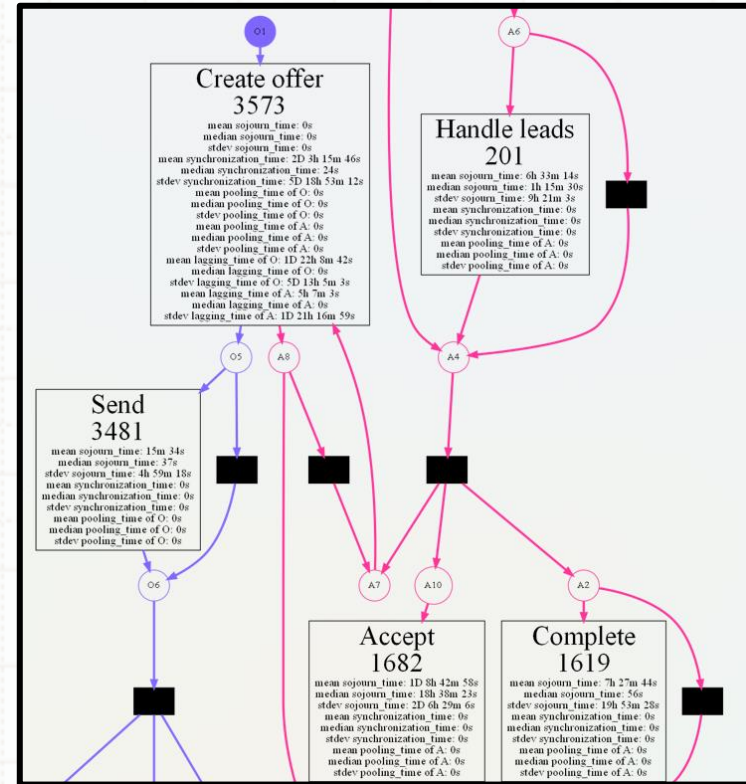


Thí nghiệm 3 - Phân tích và tinh chỉnh mô hình quy trình của phương pháp xem xét nhiều đối tượng (thí nghiệm mở rộng ngoài bài báo gốc)

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

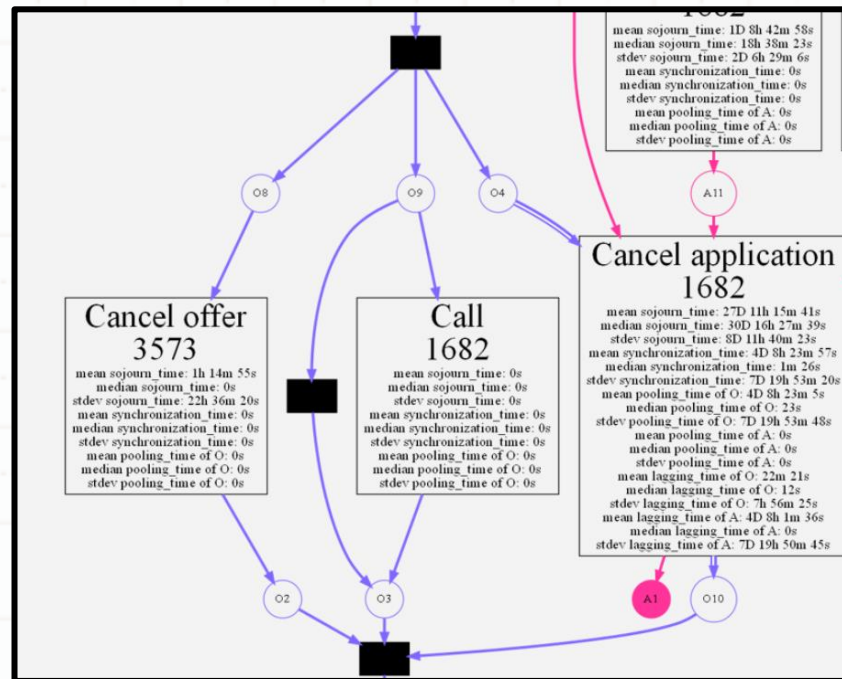
Điểm bất thường 1

- Đối với các hoạt động Accept, Complete và Create Offer, khi xét loại đối tượng là đơn đăng ký trên mô hình quy trình hiện tại ở thí nghiệm 1 đưa ra kết quả rằng ba hoạt động này **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.



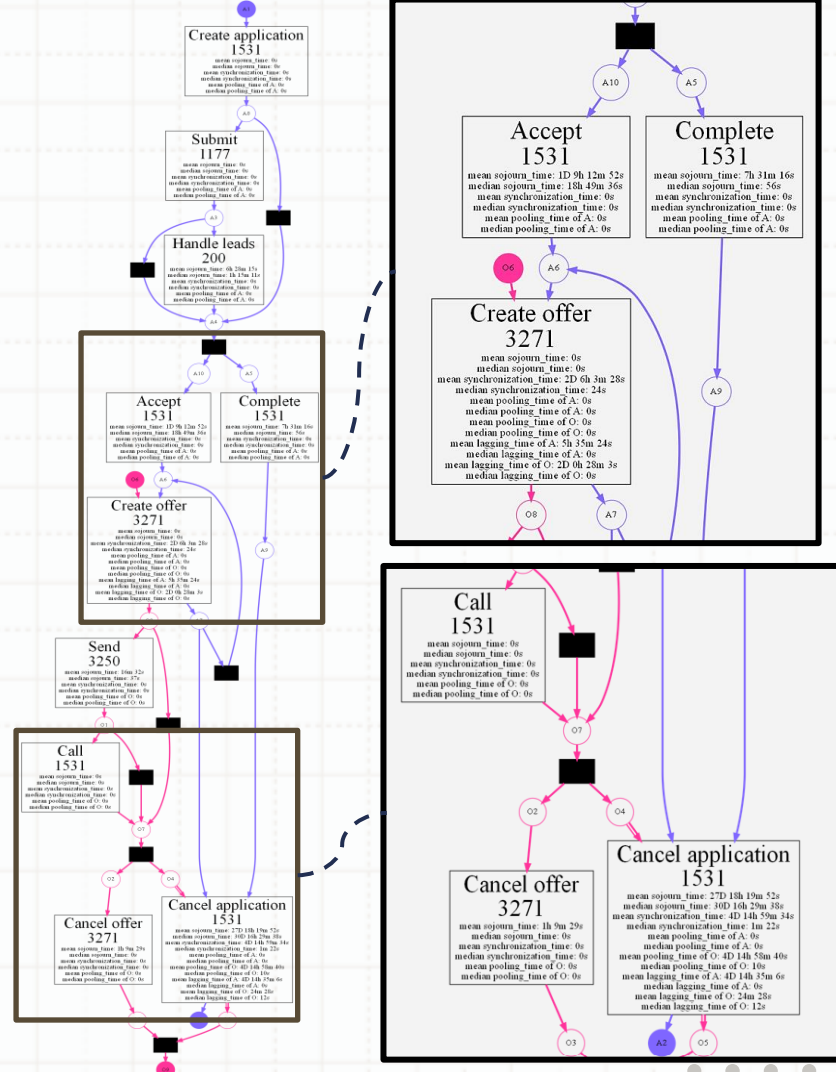
Điểm bất thường 2

- Đối với các hoạt động Cancel Offer, Cancel Application và Call khi xét với loại đối tượng đề xuất 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 loại bỏ các quy trình chưa phù hợp 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 giải quyết bài toán.

02

Phương pháp xem xét nhiều đối tượng

03

Thực nghiệm

04

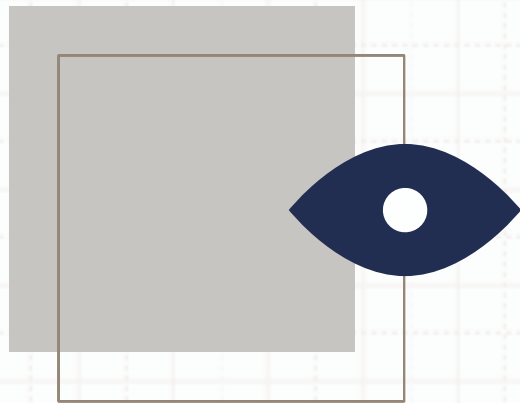
Tổng kết và hướng phát triển



04

TỔNG KẾT VÀ HƯỚNG PHÁT TRIỂN

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





Tổng kết

- Trong khóa luận này, chúng em đã tìm hiểu và trình bày một phương pháp mới trong lĩnh vực khai thác quy trình đó là 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 và có được các kết quả thí nghiệm từ cài đặt khóa luận này.
- Có được các kết quả thí nghiệm mở rộng để thấy rõ hơn về ưu/nhược điểm của phương pháp.



Hướng phát triển

- Mở rộng để hỗ trợ phân tích hiệu năng ngay cả khi dữ liệu event log không tuân thủ các quy tắc nhất định.
- Phát triển để phân tích hiệu năng tập trung vào từng đối tượng dựa trên dữ liệu event log mà không cần phụ thuộc quá nhiều vào các mô hình quy trình.
- Xây dựng và tính toán các chỉ số hiệu năng mới, thú vị hơn, xem xét sự tương tác giữa các đối tượng để cung cấp các phân tích sâu sắc và toàn diện hơn về hiệu năng của quy trình.



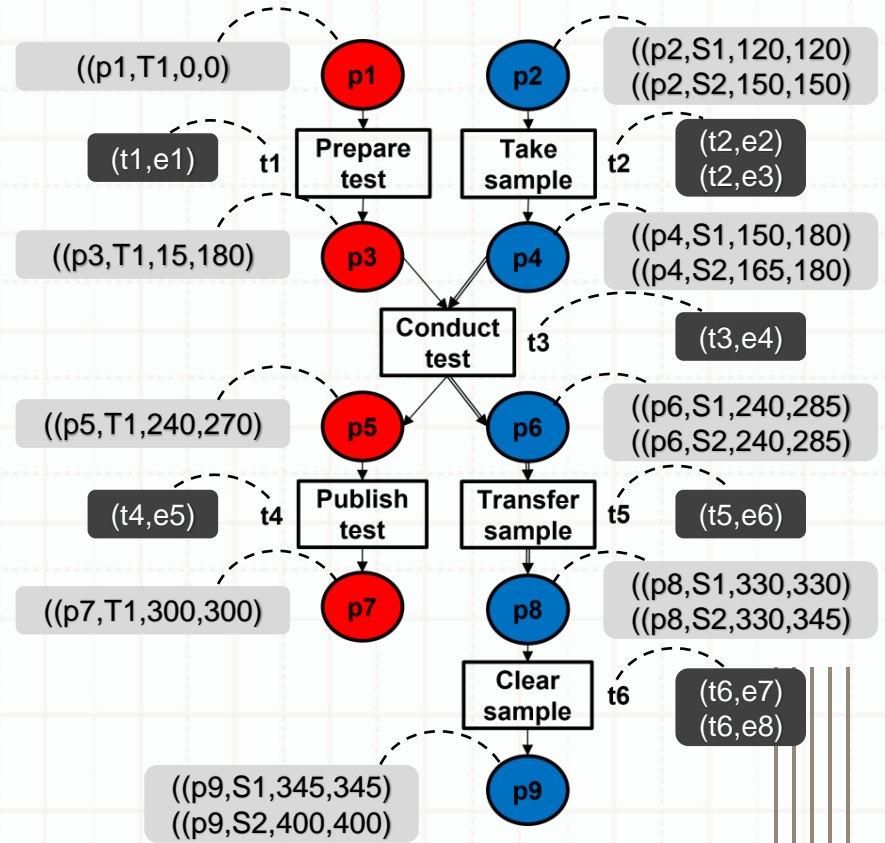
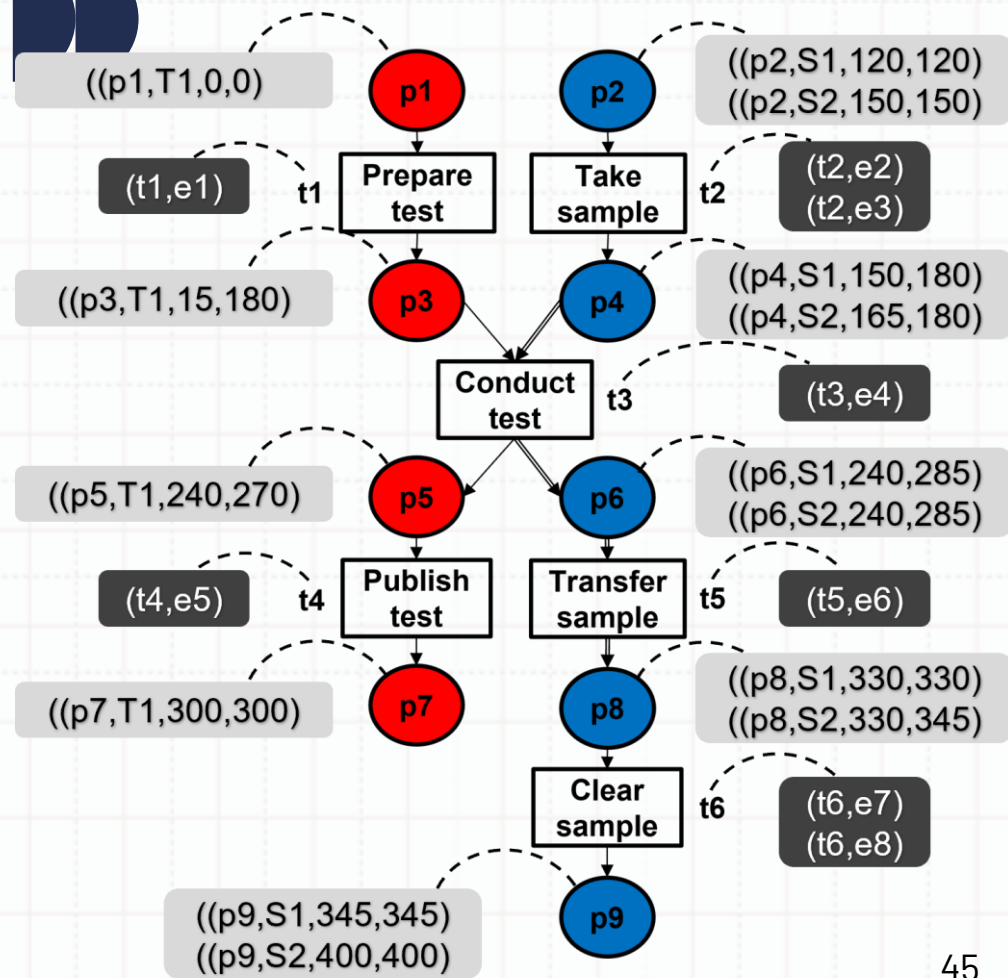
CẢM ƠN THẦY CÔ ĐÃ LẮNG NGHE





Dữ liệu đầu vào và yêu cầu của bài toán

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	165
e5	Publish test	T1		270	300
e6	Transfer sample		S1, S2	285	330
e7	Clear sample		S1	330	345
e8	Clear sample		S2	345	400





Nhận xét

Việc điều chỉnh mô hình đã giúp mô hình phản ánh chính xác hơn các chi tiết và thứ tự thực hiện của các hoạt động trong quy trình. Điều này không chỉ giúp cải thiện độ chính xác của mô hình mà còn cung cấp cái nhìn sâu sắc và chi tiết hơn về các khía cạnh hiệu năng của quy trình.

- Không có sự chồng chéo giữa các đơn đăng ký, mỗi đơn đăng ký được xử lý một cách độc lập và riêng biệt.
- Quá trình hủy các đơn đăng ký thường kéo dài đặc biệt là ở các quy trình có đơn đăng ký chưa chuẩn xác về thông tin hay quy trình có đề xuất cần có sự tư vấn hay xác nhận của khách hàng trước khi bị hủy bỏ.



Tổng kết

Phương pháp này có các ưu nút place như:

- Tính chính xác và linh hoạt: phương pháp xem xét nhiều đối tượng cho phép tính toán chính xác các chỉ số hiệu năng trong bối cảnh các quy trình phức tạp có sự liên quan đến nhiều đối tượng khác nhau.
- Hỗ trợ các độ đo hiệu năng mới: phương pháp này không chỉ giúp tính toán chính xác các chỉ số hiệu năng truyền thống mà còn hỗ trợ các chỉ số hiệu năng mới dựa trên sự tương tác giữa các đối tượng.



Tổng kết

Ngoài những ưu nút place trên, phương pháp xem xét nhiều đối tượng vẫn tồn tại những hạn chế nhất định:

- Phụ thuộc vào chất lượng của mô hình quy trình: phương pháp này phụ thuộc rất nhiều vào các mô hình quy trình được khám phá từ dữ liệu event log đầu vào.
- Chất lượng của dữ liệu đầu vào: khi các sự kiện trong dữ liệu event log bị thiếu, mất hoặc có thông tin không chính xác thì dữ liệu sẽ không phù hợp hoặc không tương thích hoàn toàn với quy trình mà mô hình đã mô tả.

Cách tiền xử lý 2

- Các hoạt động như Send, Cancel Application và Cancel Offer đã không còn xuất hiện vòng lặp. Tuy nhiên, do việc tiền xử lý này đã phân tách riêng biệt các loại đối tượng với nhau nên sẽ có đến hai mô hình quy trình được tìm thấy.
- Thời gian lưu trú trung bình cho hoạt động Send theo được tính toán là khoảng 15 phút.
- Thời gian lưu trú trung bình cho hoạt động Cancel Application cho ra hai kết quả khác nhau: một kết quả là khoảng 27 ngày 12 giờ và một kết quả khác là khoảng 29 ngày 23 giờ.

Nhìn chung, phương pháp xem xét nhiều đối tượng vẫn là phương pháp thể hiện đầy đủ và chính xác nhất những gì diễn ra trong quy trình thực tế, đặc biệt là sự tương tác giữa các đối tượng trong quy trình, làm cho nó trở thành lựa chọn tối ưu hơn so với phương pháp truyền thống.

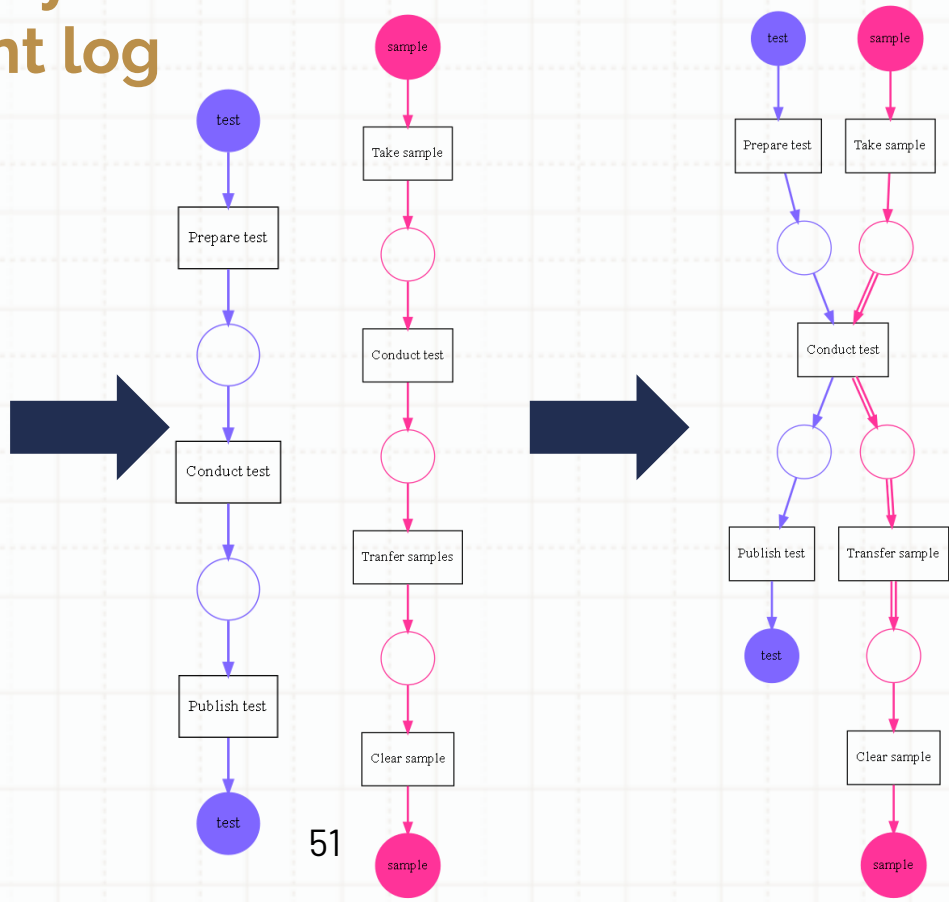
Tìm mô hình quy trình có xem xét nhiều đối tượng từ dữ liệu event log

Ý tưởng: sử dụng các nút place (places) và nút transition (transition) để mô tả các tương tác giữa các loại đối tượng và sự kiện trong quy trình. Quá trình này bao gồm xác định và phân tách các loại đối tượng, áp dụng kỹ thuật khai thác quy trình để tạo mô hình Petri Net cho từng loại, sau đó tích hợp các mô hình này thành một mô hình tổng thể (Object-Centric Petri Net), gán đối tượng cho các nút place, xác định các cung thay đổi và hoàn tất mô hình.

Kết quả: mô hình quy trình Object-Centric Petri Net (một biến thể của mô hình Petri Net), tập trung vào việc biểu diễn các quy trình dựa trên đối tượng.

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

eID	Activity	Test	Sample	Start	Complete
e1	Prepare test	T1		0:00:00	0:00:15
e2	Take sample		S1	0:02:00	0:02:30
e3	Take sample		S2	0:02:30	0:02:45
e4	Conduct test	T1	S1, S2	0:03:00	0:04:00
e5	Publish test	T1		0:04:30	0:05:00
e6	Transfer sample		S1, S2	0:04:45	0:05:30
e7	Clear sample		S1	0:05:30	0:05:45
e8	Clear sample		S2	0:05:45	0:06:40



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

Ý tưởng: kết hợp các sự kiện trong dữ liệu event log với Object-Centric Petri Net bằng cách “chơi” trò chơi token sử dụng các ngữ nghĩa hình thức từ Object-Centric Petri Net.

Kết quả:

- Tập hợp sự diễn ra của các sự kiện (event occurrences) được chú thích cho mỗi lần transition được hiển thị trên Object-Centric Petri Net.
- Tập hợp sự ghé thăm của các token (token visits) được chú thích cho mỗi nút place (place) trong mô hình.

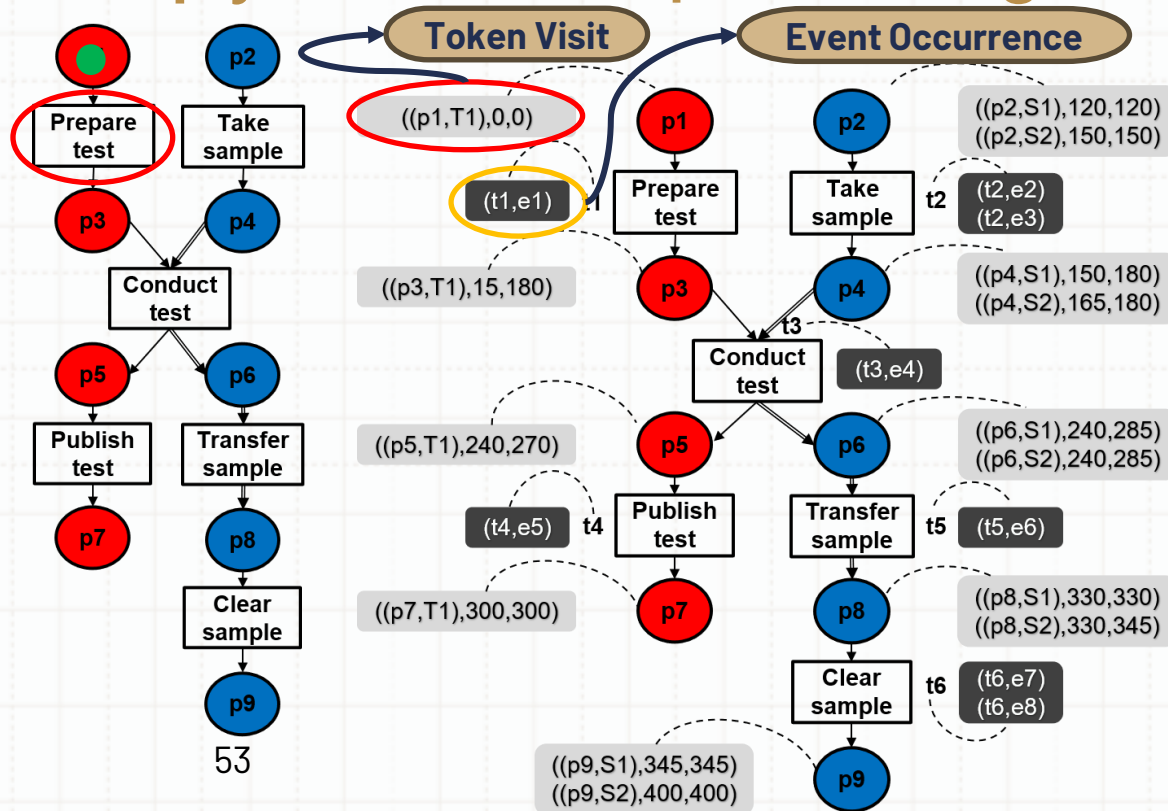
Tính các kết quả trung gian liên quan đến chỉ số hiệu năng từ mô hình quy trình và dữ liệu event log

Event Occurrence

Thể hiện sự diễn ra của một sự kiện liên quan đến một quá trình nút transition.

Token Visit

Mô tả sự ghé thăm của một token tới một nút place tương ứng.



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

Để thực hiện việc đo lường các chỉ số hiệu năng, trước tiên cần liên kết các thông tin của một sự diễn ra của sự kiện với lượt ghé thăm của các token liên quan đến sự kiện đó.

