FLOW ~~Process~~ Mining & RFID

Dennis Lin
University Distinguished Professor
Department of Supply Chain & Information Systems
The Pennsylvania State University

DennisLin@psu.edu

What is a Process?

a series of operations performed in the making or treatment of a product

a particular course of action intended to achieve a result

Process Mining

- How to Wash Your Hands?
  - Five Steps for Washing your hands
- When your company faces problems, is there any formal procedure to follow?
  - Business Process Mining
    - Via Strategy/Design
    - Via Data

How to Wash your Hands?

正確洗手五步驟

2. 取清潔劑搓洗指尖、指縫、手心、手臂，至少20秒

WET

RUB

POUR

1. 脫下手中飾物，手沾濕

3. 在流動的自來水下，沖淨雙手。

HOLD

WIPE

4. 捧水沖洗水龍頭。

5. 用乾淨的紙巾擦乾雙手，以用過的擦手紙關緊水龍頭。

Penn State (Nov 07)

Wash your hands!

1. Wet Hands
2. Soap
3. Lather for 15 sec.
4. Rinse
5. Dry
6. Turn H_2O Off

More Example: Process

- Quality Control
  - Plan-Do-Act-Check
- Six-Sigma
  - Define-Measure-Analysis-Improve-Control

- How to organize a big conference?
- How about Business Process?
  - RFID Example

Six Sigma Methodologies

Improvement Model: DMAIC

(Re)Design Model: DMADV

Develop a standard procedure/process when unexpected happens!

How to build up the process?

Is this an optimal process?

Statistical Process Mining

Data-Based Process Mining

Process Mining: Data

case identifier	task identifier
case 1	task A
case 2	task A
case 3	task A
case 3	task B
case 1	task B
case 1	task C
case 2	task C
case 4	task A
case 2	task B
case 2	task D
case 5	task E
case 4	task C
case 1	task D
case 3	task C
case 3	task D
case 4	task B
case 5	task F
case 4	task D

Table 1. A process log.

Process Minina: Model

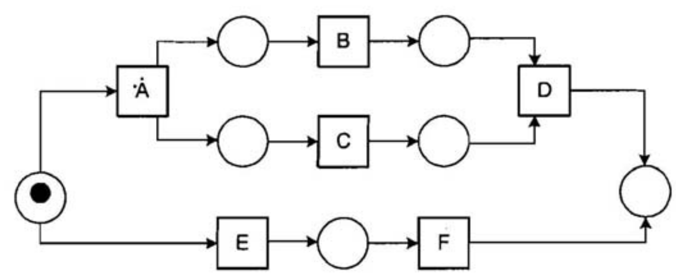
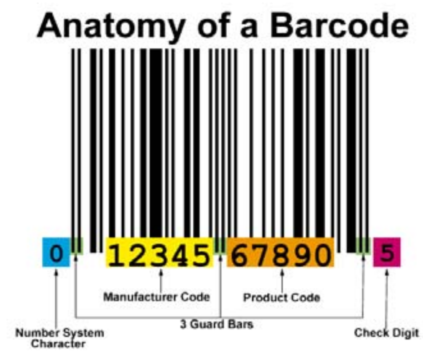


Fig. 1. A process model corresponding to the process log.

RFID:

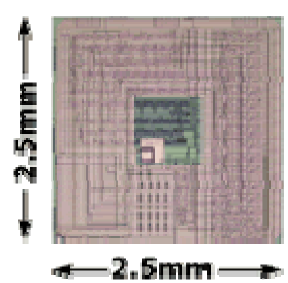
Radio Frequency IDentification

Bar Code



RFID:

Radio Frequency Identification



The components of a 96-bit electronic product code (in Hex)

01	0000ABC	000123	000056789
Header 8 bits	EPC Manager (will be assigned to companies) 28 bits	Object Class - Product Code 24 bits	Serial Number 36 bits

The RFID tag responds to the reader by broadcasting its EPC, which is a 96-bit code consisting of:

· 8 bits of header information

· 28 bits identifying the organization that assigned the code

· 24 bits identifying the type of product

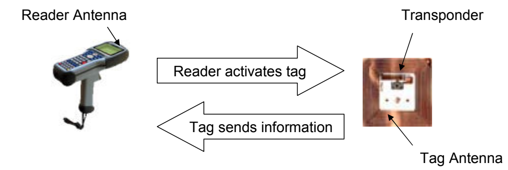
· 36 bits representing serialization information for the product

Source: Avicon white paper, 2003

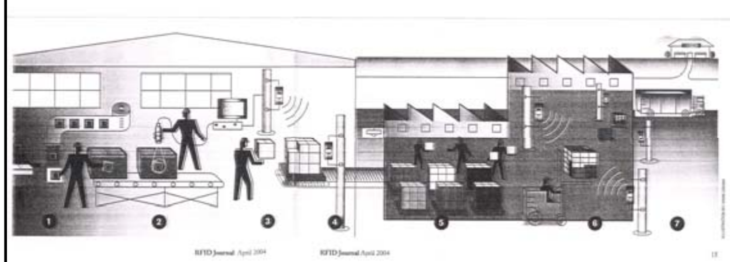
RF Communication

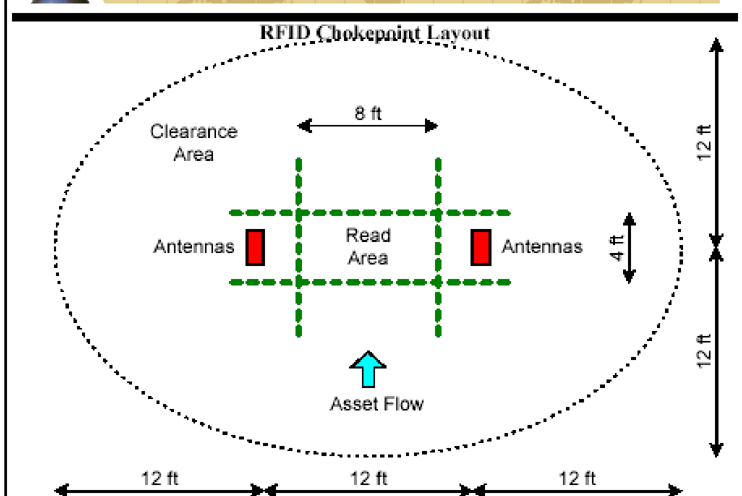
- Electromagnetic waves modulated to carry data/signals
- Two different ways to generate ways
  - Inductive coupling
    - Close proximity electromagnetic wave
  - Propagating electromagnetic waves
- The fundamental RF communication theories apply—nothing new.
- New: the cost, size, signal processing capability.

RFID

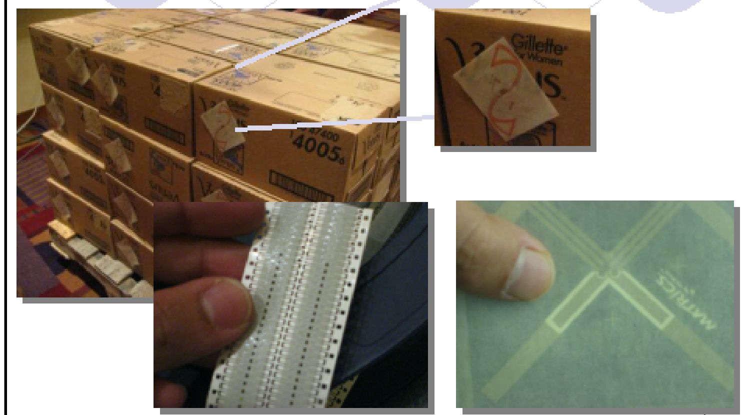


- Radio Frequency Identification (RFID) technology has been in use since the 1950's
- Advocated as Electronic Product Code™ or EPC™





RF-ID Tags



Per Mr. Jun Takei

Low Frequencies *High Frequencies*

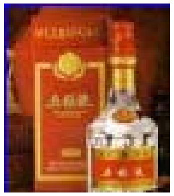
Some Current Applications

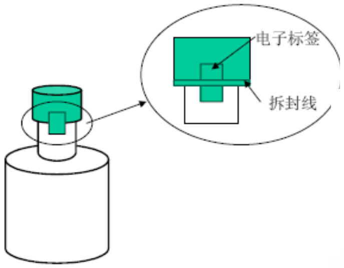
Auto
Theft Protection

134.2kHz Low Frequency

Waste Management
(Shanghai, China)

Illegal Liquor Protection



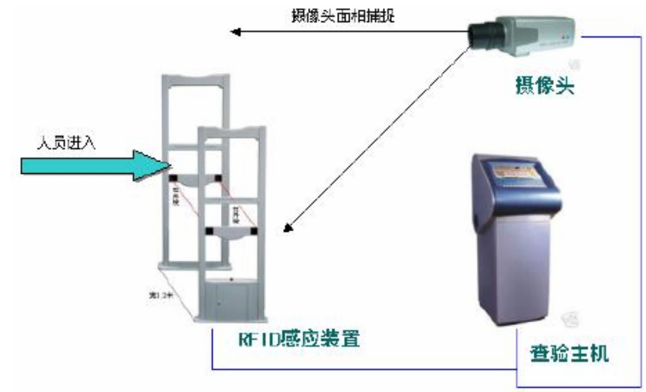


134.2kHz Low Frequency

Hotel
Management

Entrance Control

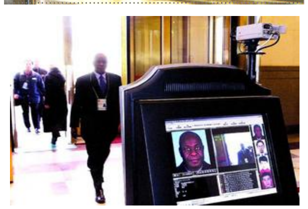
Security Management



13.56MHz High Frequency

Security Management

- APEC
- Olympic at Beijing
- International Meetings
- People & Cars



13.56MHz High Frequency

Reformation (Jail) Management

Hospital Management

Japan school kids to be tagged with RFID chips

The chips will be put onto kids' schoolbags, name tags or clothing to track the kids' movements.

By Jo Best

27 March, 2006

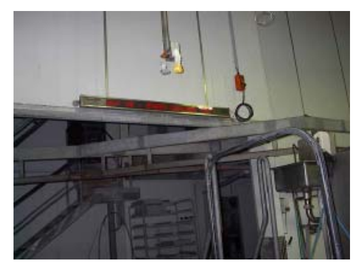
http://news.com.com/Japan+school+kids+to+be+tagged+with+RFID+chips/2100-1012_3-5266700.htm?tag=sas.email

Library Book Management (SNG)

Application Possibilities

- Shelf Reading
- Shelving
- Inventory Management
- Data Collection
- Determining Shelf Order
- Searching
- Finding Lost/Missing Items
- Pre-sort Activities
- Weeding Items

RFID in Animal Husbandry



Next Generation RFID!!

EPC

- LF: 100-150 KHz

  Hotel Card, Auto Protection, Manufacturing Line, Animal Husbandry
- HF: 13.56 MHz
  - 14443
    - 14443A—CashCard
    - 14443B—ID, Passport
  - 15693—10cm with password
- UHF (1m)
  - 915MHz: EPC ISo1800-6C, Inventory & Supply Chain
  - 2.45 GHz
  - 5.8 GHz: High Speed EZ-Pass

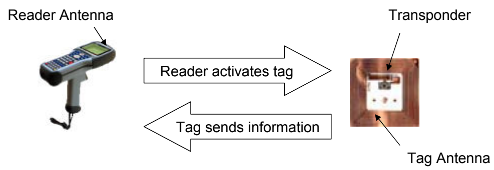
A Data Explosion is Coming!

Are You Ready?

Impact to Statistics

How to analyze the population data?

RFID



Reader ***Information System*** **Tag**

RFID vs. Barcode

Lin, Dennis K.J. and Wadhwa, Vijay
(Cover Story of Quality Progress, Feb 2007)

Quality Progress (February 2008)

SUPPLY CHAIN MANAGEMENT

by Dennis Lin and Vijay Wadhwa

Efficiency GETS A NEW IDENTITY

Barcode and RFID in the supply chain

AUTOMATED DATA COLLECTION systems (auto-ID) consist of many technologies, including barcode, voice systems, radio frequency identification (RFID), pick to light and laser scanners.[1] These systems allow for noncontact reading and are effective in a wide range of applications.

The most commonly used auto-ID technique is barcode, but RFID is catching up, primarily because it has been adopted by many of the world's largest retailers, including Wal-Mart, Target and Tesco. This article demonstrates the impact RFID will have on existing supply chain processes and the improvements RFID implementation will bring by comparing the benefits of RFID with those of barcode for various supply chain entities.

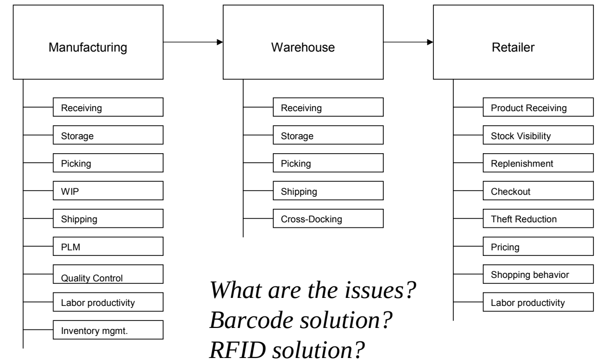
Bar-coding

- Bar coding is one of the most popular and cost effective forms of automated data collection systems.
- Fixed Barcode: Contains static information that is the same for all products of the same brand and type. An example is a UPC bar code on a 12 oz. can of coke.
- Variable barcode: A variable data bar code contains data that identifies a single product and changes for each separate product.

RFID

- RFID uses radio waves to automatically identify people or objects.
- Old technology but increased affordability, scalability, data processing capability.
- Capable of identifying each and every object uniquely in a supply chain.

Basic Structure



What are the issues?
Barcode solution?
RFID solution?
Comparisons!!!

Data Structure: S.I.T. Space

- State
- ID
- TimeStamp

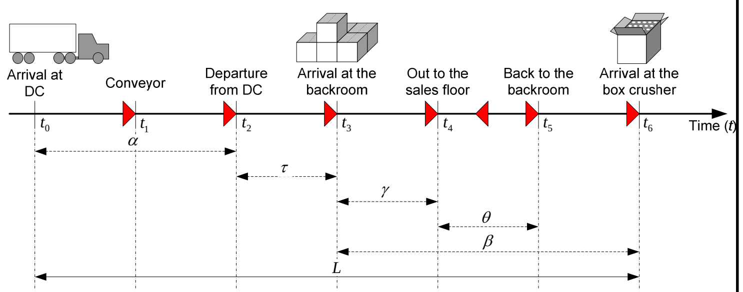
- Namely, {where, which, when}

- Particularly interested in the difference between *Execution* & *Planning* *(Sense & Response)*

Sample RFID Data

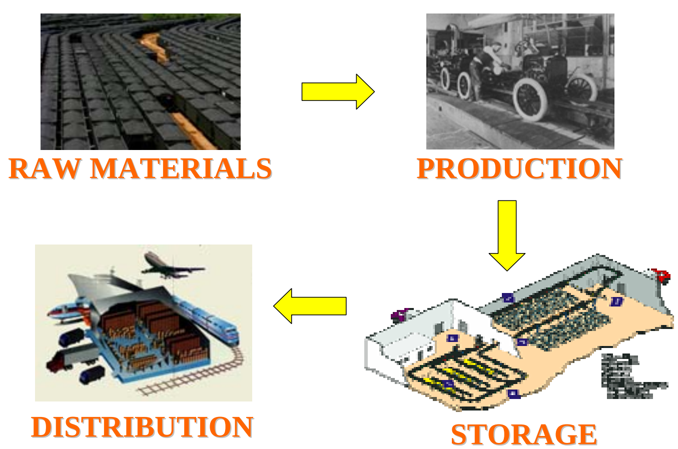
Location	EPC	Date/time	Reader
DC 123	0023800.341813.500000024	08-04-05 23:15	inbound
DC 123	0023800.341813.500000024	08-09-05 7:54	conveyor
DC 123	0023800.341813.500000024	08-09-05 8:23	outbound
ST 987	0023800.341813.500000024	08-09-05 20:31	inbound
ST 987	0023800.341813.500000024	08-09-05 20:54	sales floor
ST 987	0023800.341813.500000024	08-10-05 1:10	sales floor
ST 987	0023800.341813.500000024	08-10-05 1:12	backroom
ST 987	0023800.341813.500000024	08-11-05 15:01	sales floor
ST 987	0023800.341813.500000024	08-11-05 15:47	sales floor
ST 987	0023800.341813.500000024	08-11-05 15:49	box crusher

The timeline in a supply chain

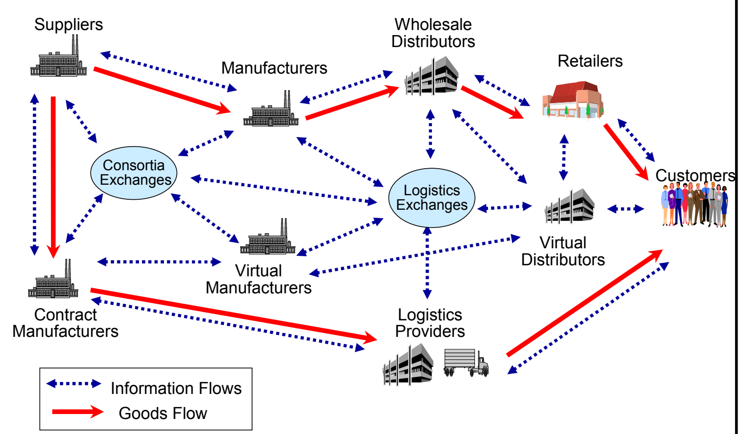


Process Mining & Flow Time Analysis

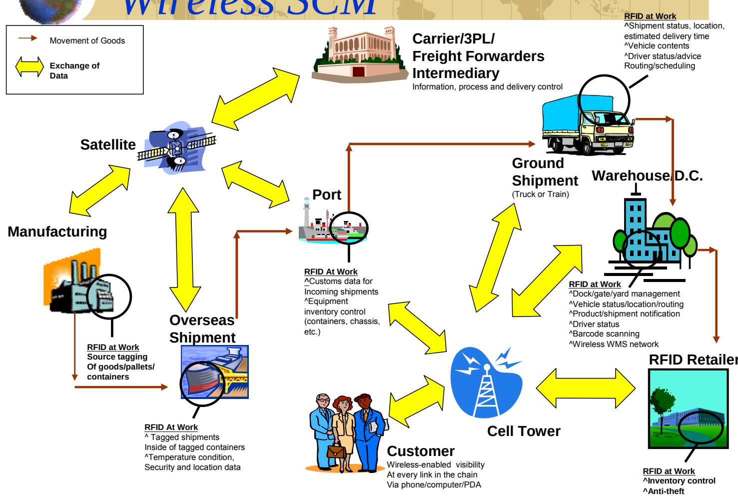
Example of a Supply Chain



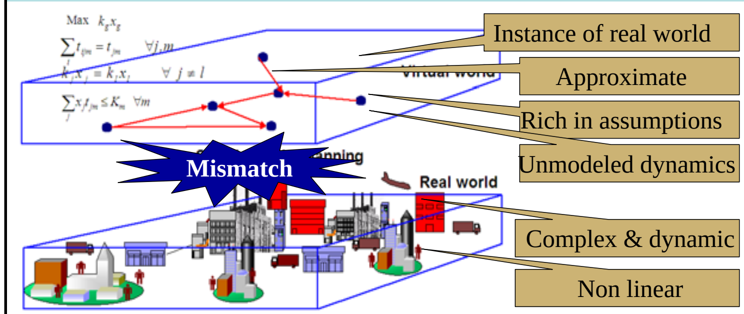
What is new? *e-Supply Chain*



Wireless SCM



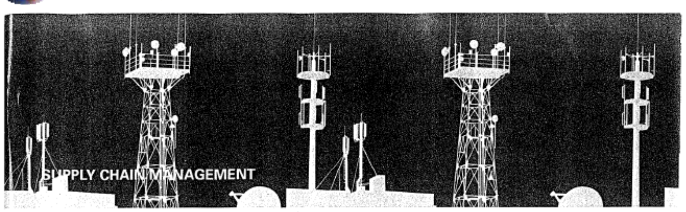
Essence of event management



SCEM accepts this discrepancy and is focused at improvements

PENNSTATE

Lin and PSU Study Group (2006)



Challenges in RFID Enabled Supply Chain Management

by RFID Study Group at Pennsylvania State University

Process Mining: Data

case identifier	task identifier
case 1	task A
case 2	task A
case 3	task A
case 3	task B
case 1	task B
case 1	task C
case 2	task C
case 4	task A
case 2	task B
case 2	task D
case 5	task E
case 4	task C
case 1	task D
case 3	task C
case 3	task D
case 4	task B
case 5	task F
case 4	task D

Table 1. A process log.

Process Mining: Model

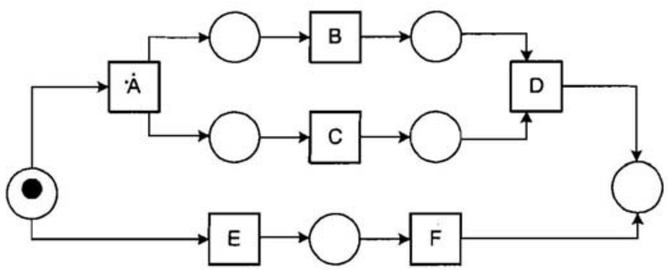
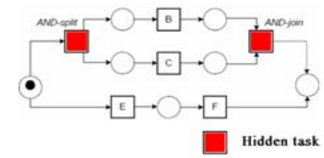


Fig. 1. A process model corresponding to the process log.

Mining Process Model- Mining Hidden Tasks

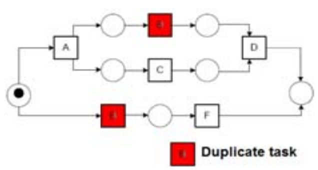
Case identifier	Task identifier
~~Case 1~~	~~Task A~~
~~Case 2~~	~~Task A~~
~~Case 3~~	~~Task A~~
Case 3	Task B
Case 1	Task B
Case 1	Task C
Case 2	Task C
~~Case 4~~	~~Task A~~
Case 2	Task B
~~Case 2~~	~~Task D~~
Case 5	Task E
Case 4	Task C
~~Case 1~~	~~Task D~~
Case 3	Task C
~~Case 3~~	~~Task D~~
Case 4	Task B
Case 5	Task F
~~Case 4~~	~~Task D~~



Even task A is removed from the log, it is clear that there has to be an AND-split if we assume tasks B and C to be in parallel. Similar for D.

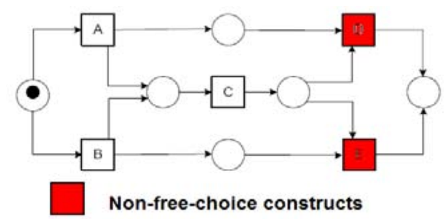
Mining Process Model- Mining Duplicate Tasks

Case identifier	Task identifier
Case 1	Task A
Case 2	Task A
Case 3	Task A
Case 3	Task B
Case 1	Task B
Case 1	Task C
Case 2	Task C
Case 4	Task A
Case 2	Task B
Case 2	Task D
Case 5	Task E → B
Case 4	Task C
Case 1	Task D
Case 3	Task C
Case 3	Task D
Case 4	Task B
Case 5	Task F
Case 4	Task D



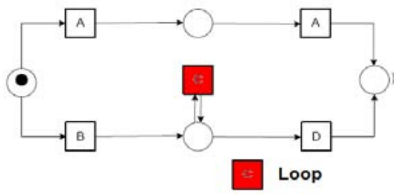
Questionable process model

Mining Process Model-Mining Non-free-choice Constructs



The choice between task D and task E is decided not only by their immediately precedent, but also some earlier choices (A and B). Such constructs are difficult to mine since the choice is non-local and the mining algorithm has to ''remember'' earlier events.

Mining Process Model-Mining loops



Mining loops can be difficult if the loops includes many tasks (long span) and involve splits/joins.

Research Issues: Theoretical & Applications

- How to build up the model (flow-chart), for noise-free cases?
- How to build up the model (flow-chart), with noise?
- How to make use of the model?
- How to compare two models? What is the “optimal” model?
- Multiple Process Mining?

Standard language

- Standard language
- Unified Modeling Language (UML)
- Event-Driven Process Chain (EPC)
- Petri Nets

Petri Nets

- Places
  - Represents one or many objects. Each object is always in some state.
- Transitions
  - Represents one or many operations, which are only possible at specific states of objects and which change the state of specific objects
- Arcs

What kind of data?

- Process mining: event log

Table 1 An event log

Event	Case ID	Activity ID	Originator	Time stamp
1	Case 1	Activity A	John	9-3-2004: 15.01
2	Case 2	Activity A	John	9-3-2004: 15.12
3	Case 3	Activity A	Sue	9-3-2004: 16.03
4	Case 3	Activity B	Carol	9-3-2004: 16.07
5	Case 1	Activity B	Mike	9-3-2004: 18.25
6	Case 1	Activity C	John	10-3-2004: 9.23

Case 1:{ABCD}
Case 2:{ACBD}
Case 3:{ABCD}
Case 4:{ACBD}
Case 5:{EF}

The resulting model

Case 1:{ABCD}
Case 2:{ACBD}
Case 3:{ABCD}
Case 4:{ACBD}
Case 5:{EF}

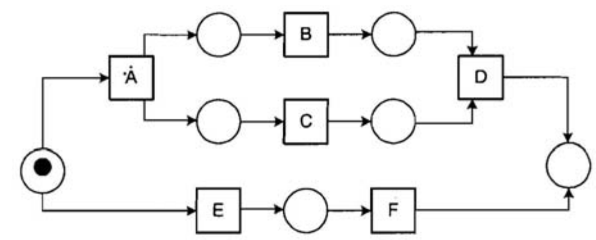


Fig. 1. A process model corresponding to the process log.

Literature review: Algorithms

- Software (most based on α-algorithm)
  - Emit (Van Aalst et. al., 2002)
  - Little Thumb (Van Aalst et. al., 2004)
  - InWoLvE (Herbst et. al., 2001)
  - Process Miner (Schimm et. al., 2002)

Difficulties and Challenges

- Mining hidden activities
- Mining duplicate activities
- Mining non-free-choice constructs
- Mining loop
- Using time
- Visualizing results
- Mining different perspectives
- Dealing with noise
- Dealing with incompleteness
- Delta analysis

Van Aalst et. al., 2004

α-algorithm

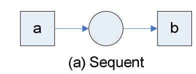
- If a and b are sequent iff a >b (a is the input of b)
- If a and b are alternative iff a $\not>$ b and b $\not>$ a

  (a can't be the input of b and b can't be the input of a)
- If a and b are concurrence iff a>b and b>a

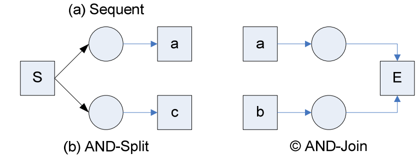
  (a can be the input of b and b can be the input of a)

α-algorithm: Basic Elements

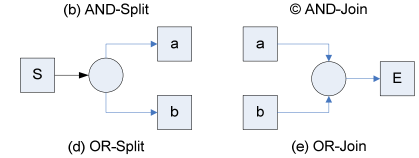
- Sequent



- Concurrence



- Alternative



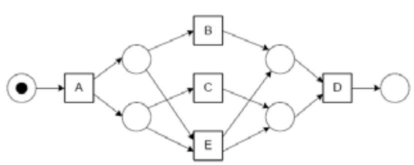
- *Loop*

Optimal Model: Illustrative Example

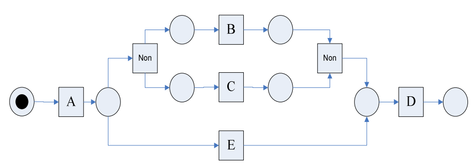
Event logs *W:*

{ABCD}, {ACBD}, {AED}

Model-1



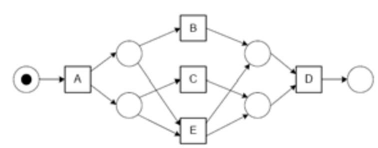
Model-2



Optimal Model: Illustrative Example

Event logs *W:*

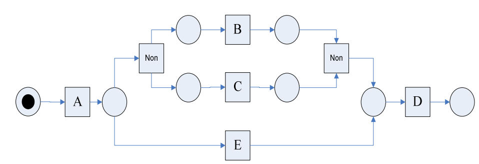
{ABCD}, {ACBD}, {AED}



{ABCD} {ACBD} {AED}

{ACED} {AECD} {ABED}

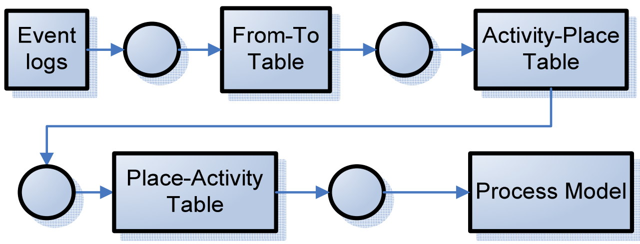
{AEBD}



{ABCD},
{ACBD},
{AED}

Mining Process Model

The Process of Building a process model:



Process Mining: Data

case identifier	task identifier
case 1	task A
case 2	task A
case 3	task A
case 3	task B
case 1	task B
case 1	task C
case 2	task C
case 4	task A
case 2	task B
case 2	task D
case 5	task E
case 4	task C
case 1	task D
case 3	task C
case 3	task D
case 4	task B
case 5	task F
case 4	task D

Table 1. A process log.

Example: Event Logs

Event logs

Case 1: SABCDEFGHI; Case 2: SACBDEFGHI;

Case 3: SAJCDEFGHI; Case 4: SACJDEFGHI;

Case 5: SABCDEFEFGHI; Case 6: SACBDEFEFGHI

Case 7: SAJCDEFEFGHI; Case 8: SACJDEFEFGHI

Case 9: SABCDEFGH ABCDEFGHI;

Case 10: SABCDEFGHACBDEFGHI;

Case 11: SACBDEFGHAJCDEFGHI;

Case12: SAJCDEFGH ABCDEFEFGHI;

Frequencies?
Time Sequence?

Step1 ：Build an n×n From-To Table

From\To	S	A	B	C	D	E	F	G	H	I	J
S	0	1	0	0	0	0	0	0	0	0	0
A	0	0	1	1	0	0	0	0	0	0	1
B	0	0	0	1	1	0	0	0	0	0	0
C	0	0	1	0	1	0	0	0	0	0	1
D	0	0	0	0	0	1	0	0	0	0	0
E	0	0	0	0	0	0	1	0	0	0	0
F	0	0	0	0	0	1	0	1	0	0	0
G	0	0	0	0	0	0	0	0	1	0	0
H	0	1	0	0	0	0	0	0	0	1	0
I	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	1	1	0	0	0	0	0	0

Step2 ：Find out the potential concurrent activities from From-To Table

{B,C}

{C, J}

{E, F}

From\To	S	A	B	C	D	E	F	G	H	I	J
S	0	1	0	0	0	0	0	0	0	0	0
A	0	0	1	1	0	0	0	0	0	0	1
B	0	0	0	1	1	0	0	0	0	0	0
C	0	0	1	0	1	0	0	0	0	0	1
D	0	0	0	0	0	1	0	0	0	0	0
E	0	0	0	0	0	0	1	0	0	0	0
F	0	0	0	0	0	1	0	1	0	0	0
G	0	0	0	0	0	0	0	0	1	0	0
H	0	1	0	0	0	0	0	0	0	1	0
I	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	1	1	0	0	0	0	0	0

{B,C},{C, J}, {E, F}

Are they

Concurrence?!

Alternative?!

Loop?!

Concurrence→ Both-And (not in order)

Alternative→ Either-Or

*Loop→ Trouble**

Step 3 ：Modify the From-To Table

From\To	S	A	B	C	D	E	F	G	H	I	J
S	0	1	0	0	0	0	0	0	0	0	0
A	0	0	1	1	0	0	0	0	0	0	1
B	0	0	0	**0**	1	0	0	0	0	0	0
C	0	0	**0**	0	1	0	0	0	0	0	**0**
D	0	0	0	0	0	1	0	0	0	0	0
E	0	0	0	0	0	0	**1**	0	0	0	0
F	0	0	0	0	0	**1**	0	1	0	0	0
G	0	0	0	0	0	0	0	0	1	0	0
H	0	1	0	0	0	0	0	0	0	1	0
I	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	**0**	1	0	0	0	0	0	0

Concurrence

{B,C}

{C, J}

{E, F}

Simple Loop

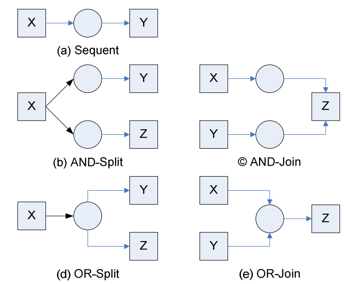
Step 4 ：Count the number of the place

{B,C,J}

Concurrence!? Alternative?!

1. {B,C} are concurrence
2. {C,J} are concurrence
3. B and J didn't happen together in any trace

Sequent? Concurrence!? Alternative?!



Step 4 ：Count the number of the place

S→A　　G→H
A→{B}{C}　　H→{I, A}
A→{J}{C}　　I→None
{B}{C}→D
{J}{C}→D
D→E
E→F
F→{E,G}

(1,named S0)

S→A	(1, named S1)
A→{B,J}{C}	(2, named S2, S3)
{B,J},{C}→D	(2, named S4, S5)
D→E	(1, named S6)
E→F	(1, named S7)
F→{E,G}	(1, named S8)
G→H	(1, named S9)
H→{I,A}	(1, named S10)
I→None	(1, named S11)

Step 5：build an *n×s* Activity-Place Table

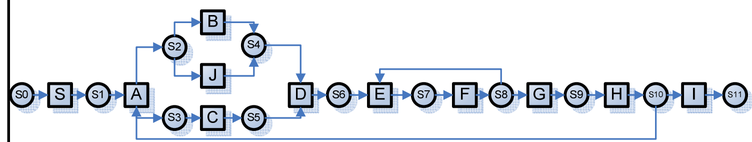
From\to	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
S	0	1	0	0	0	0	0	0	0	0	0	0
A	0	0	1	1	0	0	0	0	0	0	0	0
B	0	0	0	0	1	0	0	0	0	0	0	0
C	0	0	0	0	0	1	0	0	0	0	0	0
D	0	0	0	0	0	0	1	0	0	0	0	0
E	0	0	0	0	0	0	0	1	0	0	0	0
F	0	0	0	0	0	0	0	0	1	0	0	0
G	0	0	0	0	0	0	0	0	0	1	0	0
H	0	0	0	0	0	0	0	0	0	0	1	0
I	0	0	0	0	0	0	0	0	0	0	0	1
J	0	0	0	0	1	0	0	0	0	0	0	0

Step 6：Build an $s \times n$ Place-Activity Table

From\to	S	A	B	C	D	E	F	G	H	I	J
S0	1	0	0	0	0	0	0	0	0	0	0
S1	0	1	0	0	0	0	0	0	0	0	0
S2	0	0	1	0	0	0	0	0	0	0	1
S3	0	0	0	1	0	0	0	0	0	0	0
S4	0	0	0	0	1	0	0	0	0	0	0
S5	0	0	0	0	1	0	0	0	0	0	0
S6	0	0	0	0	0	1	0	0	0	0	0
S7	0	0	0	0	0	0	1	0	0	0	0
S8	0	0	0	0	0	1	0	1	0	0	0
S9	0	0	0	0	0	0	0	0	1	0	0
S10	0	1	0	0	0	0	0	0	0	1	0
S11	0	0	0	0	0	0	0	0	0	0	0

Step 7：

we now can build up the process model, from Activity-Place Table and Place-Activity Table,



Contribution

Given events log (data), we are able to construct a process model.

But,

Not sure whether this is the optimal one!

Comparisons with Existing Algorithms

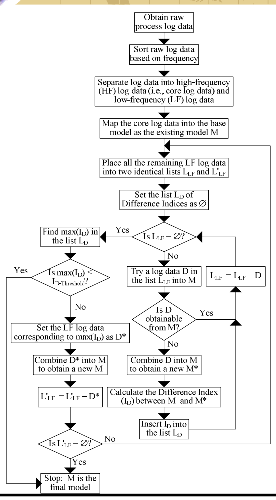
	EMit	Little Thumb	InWoLvE	Process Miner	**Proposed**
Structure	Graph	Graph	Graph	Block	**Graph**
Time	Yes	No	No	No	**No**
Basic parallelism	Yes	Yes	Yes	Yes	**Yes**
Non-free choice	No	No	No	No	**No**
Basic loops	Yes	Yes	Yes	Yes	**Yes**
Arbitrary loops	Yes	Yes	No	No	**Yes**
Hidden tasks	No	No	No	No	**Yes**
Duplicate activities	No	No	Yes	No	**No**
Noise	No	Yes	Yes	No	**Next**

Summary: Process Mining

- What is available?
- What need to be done?
- Some initial results on
  - Simple Loop
  - Non-Free Choice
  - Over-General
  - Others

Future Work

- How to build a minimum process model?
- How to build the suitable model, when the event logs data were contaminated with noise?
- How to accommodate the time (and other) component?
- When the event logs come from two (mixed) processes,
  how to build the two process models?



Step	An Algorithm of Process Mining
1	Obtain raw process log data
2	Sort raw log data based on frequency
3	Separate log data into high-frequency (HF) log data (i.e., core log data) and low-frequency (LF) log data
4	Map the core log data into the base model as the existing model M
5	Place all the remaining LF log data into two identical lists L_{LF} and L'_{LF}
6	Set of the list L_D of Difference Indices as ∅
7	If $L_{LF} \neq \emptyset$, then:
8	Try a log data D in the list L_{LF} into M
9	If D is obtainable from M, then:
10	$L_{LF} = L_{LF} - D$, and go to Step 7
11	Otherwise:
12	Combine D into M to obtain a new M*
13	Calculate the Difference Index (I_D) between M and M*
14	Insert I_D into the list L_D
15	$L_{LF} = L_{LF} - D$, and go to Step 7
16	Otherwise:
17	Find max(I_D) in the list L_D
18	If max(I_D) ≥ $I_{D\text{-Threshold}}$, then:
19	Set the LF log data corresponding to max(I_D) as D*
20	Combine D* into M to obtain a new M
21	$L'_{LF} = L'_{LF} - D^*$
22	If $L'_{LF} \neq \emptyset$, then:
23	Go to Step 5
24	Otherwise:
25	Stop: M is the final model
26	Otherwise:
27	Stop: M is the final model

Reference

- AKA de Medeiros, A.Guzzo, G.Greco, W.M.P. van der Aalst, A.J.M.M. Weijters, B.van Dongen, and D. Sacca, *Process Mining Based on Clustering: A quest for precision.* Volume 4102 of Lecture Notes in Computer Science, pages 129-144. 2008
- J. Desel and J. Esparza. *Free Choice Petri Nets*, volume 40 of Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, UK, 1995.
- LijieWen · Wil M. P. van der Aalst · Jianmin Wang · Jiaguang Sun, *Mining process models with non-free-choice constructs*, Data Min Knowl Disc (2007) 15:145–180
- References Therein