

#### Federal University of Amazonas

Institute of Computing Graduate Program in Informatics Interest Group on Embedded Systems

#### EMBEDDED MINING CORRELATIONS BETWEEN DISTRIBUTED DEVICES

Prof. Dr. Raimundo Barreto Márcio André da Costa Alencar

Salvador-BA, November 07<sup>th</sup> 2018

# PRESENTATION SCHEDULE



INTRODUCTION		Contextualization
PROBLEMATIC	II	Definition of main and specifics goals
PROPOSED METHOD	III	Components specifications and its interactions
EXPERIMENTS	IV	Experiments and results
FINAL CONSIDERATIONS	V	Important notes



#### Contextualization









V



#### Contextualization

 What is "Internet of Things"? There's no global definition but, a simplistic concept, could be as the scenario where things are connected!













#### Contextualization

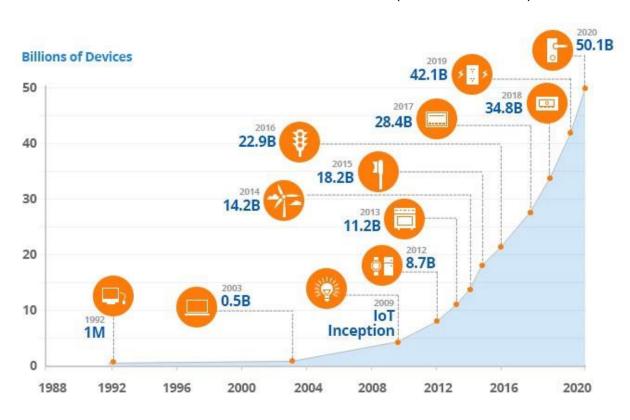
- Until 2020 will be over 50 Bi devices (IDC - 2011)



Ш

IV

V





#### Contextualization

- Many of them will be in your home;





#### Contextualization

- Many of them will be in your home;
- Be connected is not enough. IoT demands intelligence!











How to provide intelligence to IoT devices?









V

П



### How to provide intelligence to IoT devices?

The ability to provide intelligence and autonomy in IoT relies primarily on the need to identify activities, patterns and correlations implicit in the activity records of each device that composes it.

П



#### How to provide intelligence to IoT devices?

The ability to provide intelligence and autonomy in IoT relies primarily on the need to identify activities, patterns and correlations implicit in the activity records of each device that composes it.

Some points must be considered before that:

- Architecture: Centralized / Decentralized
- **Kind of intelligence**: What information extract
- **Devices Limitations**: Much / Few Resources
- Cost/Benefit: Expensive or cheap solution



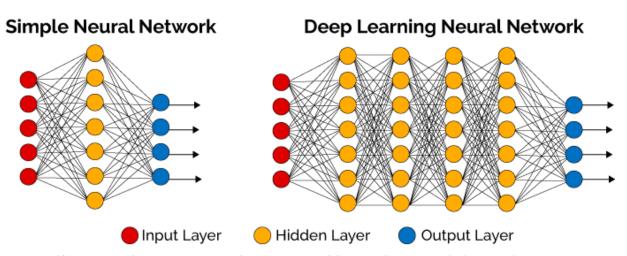
### State of art in Machine Learning

Neural Network(Deep learning)

П

IV

V

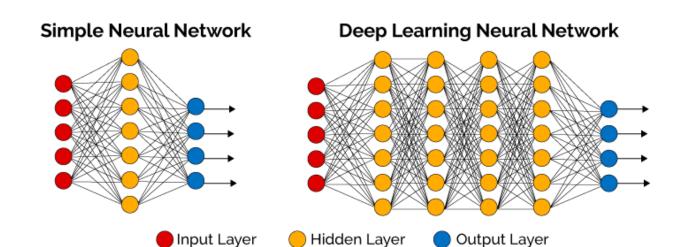




### State of art in Machine Learning

- Neural Network(Deep learning)
- High computational cost (storage/processing)

(CHEN et al., 2015; MOONS e VERHELST, 2017)



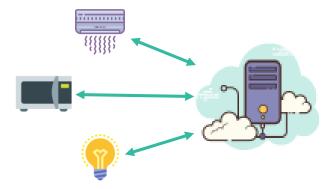


### State of art in Machine Learning

- Neural Network(Deep learning)
- High computational cost (storage/processing)

(CHEN et al., 2015; MOONS e VERHELST, 2017)

Centralization / Dependency









13



### State of art in Machine Learning

- Neural Network(Deep learning)
- High computational cost (storage/processing)

(CHEN et al., 2015; MOONS e VERHELST, 2017)

Centralization / Dependency











### State of art in Machine Learning

- Neural Network(Deep learning)
- High computational cost (storage/processing)

(CHEN et al., 2015; MOONS e VERHELST, 2017)

Centralization / Dependency









15



### State of art in Machine Learning

- Neural Network(Deep learning)
- High computational cost (storage/processing)

(CHEN et al., 2015; MOONS e VERHELST, 2017)

- Centralization / Dependency
- Devices with many resources (expensive / waste of resource)



П

IV

V



### How to provide intelligence to IoT devices...

#### Considering:

- Decentralized environment
- Devices autonomy
- Limited processing and storage resources
- Extract knowledge from limited amount of data





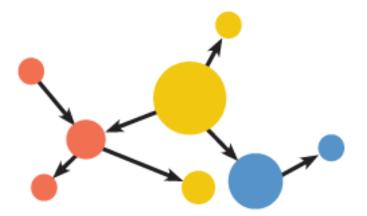




### Data mining!

A strategy to circumvent such adversities is the use of data mining techniques, but specifically of **association rules**, which seek to identify frequent patterns in a dataset so that it meets minimum support and confidence criteria.

(CHEN et al., 2015; TAN et al., 2006)











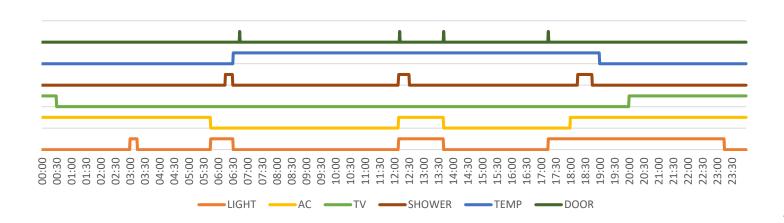


### Data mining!

П

Although each device has a unique usage pattern, some devices have similar characteristics of use, that could be used to correlate them.

(PAL et al., 2017; GONZALES e AMFT, 2015; CHEN et al., 2012; HEIERMAN et al., 2003).





#### Main Goal

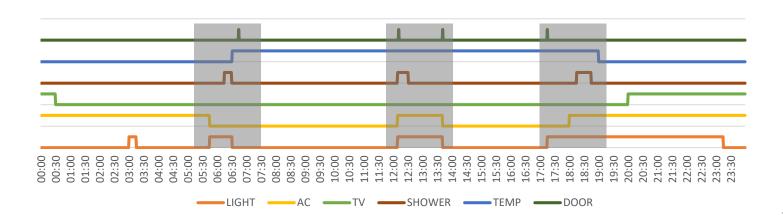
Demonstrate the ability of intelligent integration between devices in the internet of things, through a decentralized mining of implicit correlations between their **patterns of state changes**.



П

IV







### Preliminary information

It is important to comprehend some components and specifications before a general architecture overview





#### Preliminary information

It is important to comprehend some components and specifications before a general architecture overview

#### Keep that in mind:

Ш

 We want to work with devices that have severe restrictions of processing, storage and memory (Like: ESP8266-01)



#### **Preliminary information**

It is important to comprehend some components and specifications before a general architecture overview

#### Keep that in mind:

Ш

- We want to work with devices that have severe restrictions of processing, storage and memory (Like: ESP8266-01)
- All storage and processing must be performed into each device, there's no storage server or cloud services.



#### Preliminary information

It is important to comprehend some components and specifications before a general architecture overview

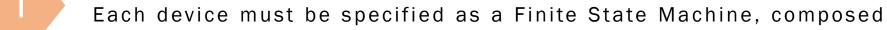
#### Keep that in mind:

Ш

- We want to work with devices that have severe restrictions of processing, storage and memory (Like: ESP8266-01)
- All storage and processing must be performed into each device, there's no storage server or cloud services.
- The devices must talk to each other (http requests) and obtain a global knowledge based on local decisions.



#### **Devices** behavior



by a set of well defined (discrete) states ( $S = \{s_1, s_2, \dots s_i\}$ ) which

represents its interactions with environment. It also has a set of

actions (  $A=\{a_1,a_2,...,a_i\}$  ) which allow to transit among the |S|

states;

Ш



#### **Devices** behavior



Ш

Each device must be specified as a Finite State Machine, composed by a set of well defined (discrete) states ( $S = \{s_1, s_2, ... s_i\}$ ) which represents its interactions with environment. It also has a set of actions ( $A = \{a_1, a_2, ..., a_i\}$ ) which allow to transit among the |S| states;

Example: Lamp/Light/Binary Device







#### Devices behavior









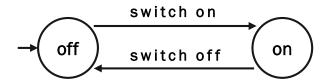




Each device must be specified as a Finite State Machine, composed by a set of well defined (discrete) states ( $S = \{s_1, s_2, ... s_i\}$ ) which represents its interactions with environment. It also has a set of actions  $(A = \{a_1, a_2, ..., a_i\})$  which allow to transit among the |S|states;

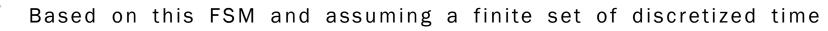
Example: Lamp/Light/Binary Device

A={"switch on","switch off"}





#### Embedded data storage



interval(slots) as  $T = \{t_1, t_2, ... t_i\}$  it is possible to define a embedded

database as a counters matrix  $M_{ij} = A imes T$  where each element  $c_{nm} \in$ 

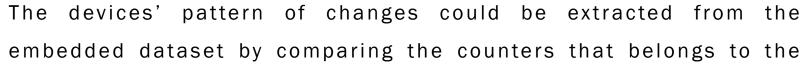
 $M_{ij}$  is a counter associated to the action  $a_n \in A$  at the time  $t_m \in T$ ;

$$M_{ij} = \begin{bmatrix} c_{11} & \cdots & c_{1j} \\ \vdots & \ddots & \vdots \\ c_{i1} & \cdots & c_{ij} \end{bmatrix}$$

$$M_{ij} = \begin{bmatrix} c_{11} & \cdots & c_{1j} \\ \vdots & \ddots & \vdots \\ c_{i1} & \cdots & c_{ij} \end{bmatrix} \quad \begin{array}{l} i: num \ of \ items \ in \ A \ (|A|); \\ j: num \ of \ items \ in \ T \ (|T|); \\ c_{nm}: counter \ for \ each \ action \ a_n \ at \ slot \ t_m; \end{array}$$



### Pattern of changes



same slots(column in matrix).

Embedded

Database(M<sub>ii</sub>)

	00:00	01:00	02:00	 21:00	22:00	23:00
switch off	10	2	0	26	14	12
switch on	2	2	0	34	17	7



#### Pattern of changes

The devices' pattern of changes could be extracted from the embedded dataset by comparing the counters that belongs to the

same slots(column in matrix).



	00:00	01:00	02:00	 21:00	22:00	23:00
switch off	<u>10</u>	2	0	26	14	<u>12</u>
switch on	2	2	0	<u>34</u>	<u>17</u>	7

Pattern of change (P)

00:00	01:00	02:00		21:00	22:00	23:00			
SWITCH OFF	-	-		SWITCH ON	SWITCH ON	SWITCH OFF			



#### Mining correlations

patterns are compared slot by slot.

Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions by the frequency of a pair of actions that shows together when two



#### Mining correlations

Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions by the frequency of a pair of actions that shows together when two patterns are compared slot by slot.

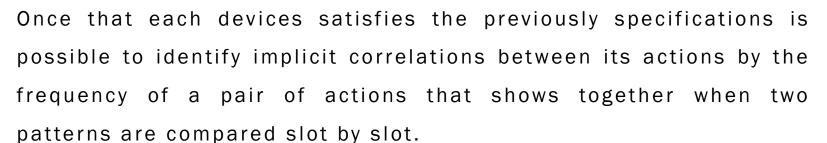
Pattern of change form Device 01

00:00	01:00	02:00	 21:00	22:00	23:00
SWITCH OFF	-	-	 SWITCH ON	SWITCH ON	SWITCH OFF





#### Mining correlations



Pattern of change form Device 01

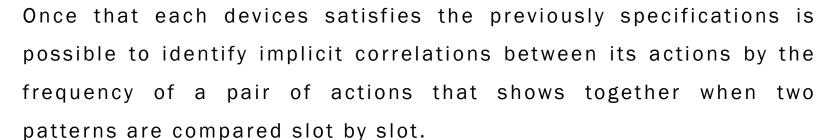
00:00	01:00	02:00	 21:00	22:00	23:00
SWITCH OFF	-	-	 SWITCH ON	SWITCH ON	SWITCH OFF

Pattern of change form Device 02

00:00	01:00	02:00	21:00	22:00	23:00
DOOR CLOSE	DOOR CLOSE	-	 DOOR OPEN	DOOR CLOSE	DOOR CLOSE



#### Mining correlations



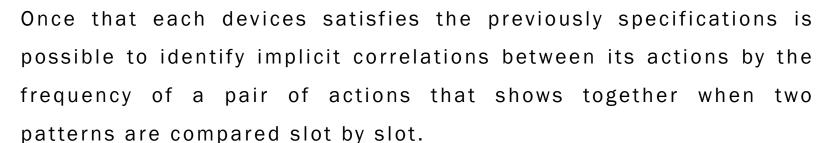




00:00	01:00	02:00	 21:00	22:00	23:00
SWITCH OFF	-	-	 SWITCH ON	SWITCH ON	SWITCH OFF
DOOR CLOSE	DOOR CLOSE	-	 DOOR OPEN	DOOR CLOSE	DOOR CLOSE



#### Mining correlations





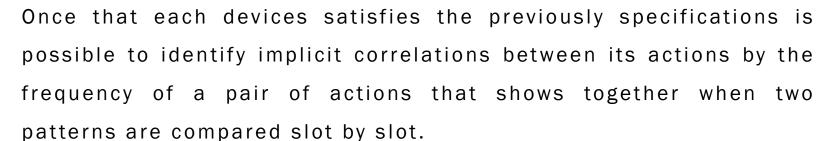
Transaction Dataset(D)

00:00	01:00	02:00	 21:00	22:00	23:00
SWITCH OFF	-	-	 SWITCH ON	SWITCH ON	SWITCH OFF
DOOR CLOSE	DOOR CLOSE	-	 DOOR OPEN	DOOR CLOSE	DOOR CLOSE

Possible Rules SWITCH OFF ⇒ DOOR CLOSE SWITCH OFF ⇒ DOOR OPEN SWITCH ON ⇒ DOOR CLOSE SWITCH ON ⇒ DOOR OPEN



#### Mining correlations





Transaction Dataset(D)

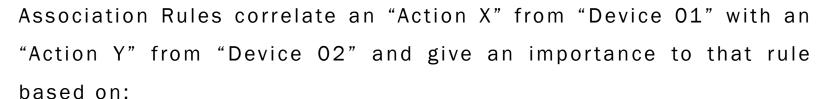
00:00	01:00	02:00	 21:00	22:00	23:00
SWITCH OFF	-	-	 SWITCH ON	SWITCH ON	SWITCH OFF
DOOR CLOSE	DOOR CLOSE	-	 DOOR OPEN	DOOR CLOSE	DOOR CLOSE

Possible Rules SWITCH OFF ⇒ DOOR CLOSE SWITCH OFF ⇒ DOOR OPEN SWITCH ON ⇒ DOOR CLOSE SWITCH ON ⇒ DOOR OPEN

Which rule(s) is(are) more frequent?



#### Mining correlations



Ш

IV

V

Device  $01(Action X) \Rightarrow Device 02(Action Y)$ 



#### Mining correlations

- Association Rules correlate an "Action X" from "Device 01" with an "Action Y" from "Device 02" and give an importance to that rule based on:
- Support: How frequent is this rule in transactions dataset(D);
- Confidence: How reliable is this rule.
- Lift: How dependent are antecedent and consequent;

Device  $01(Action X) \Rightarrow Device 02(Action Y)$ 

[ *sup*: 80%, *lift*: 4.3, *conf*: 96%]

Ш



#### Mining correlations

Put in other words, those metrics are:

Ш

Ш

IV

V

$$Support = \frac{frq(X,Y)}{N}$$

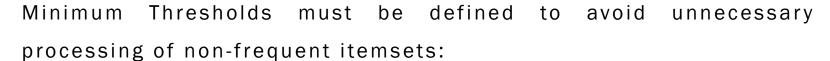
$$Rule: X \Rightarrow Y \longrightarrow Confidence = \frac{frq(X,Y)}{frq(X)}$$

$$Lift = \frac{Support}{Supp(X) \times Supp(Y)}$$

<sup>\*</sup>N = Size of dataset transaction



### Mining correlations

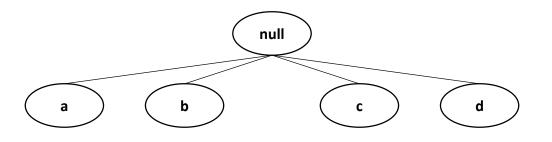






### Mining correlations

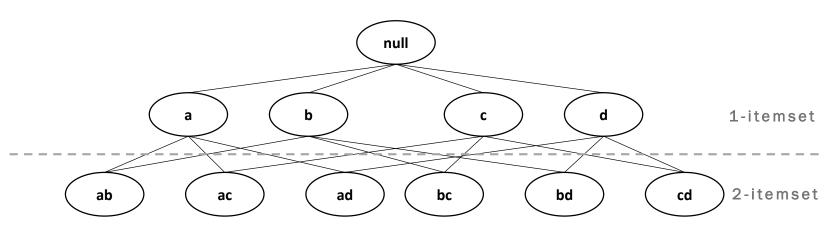
Minimum Thresholds must be defined to avoid unnecessary processing of non-frequent itemsets:



1-itemset

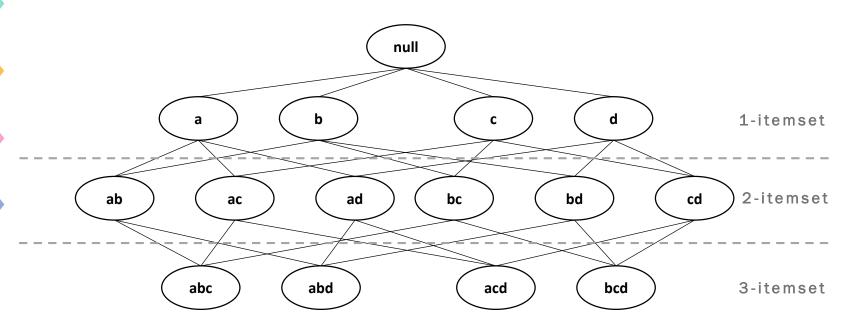


### Mining correlations



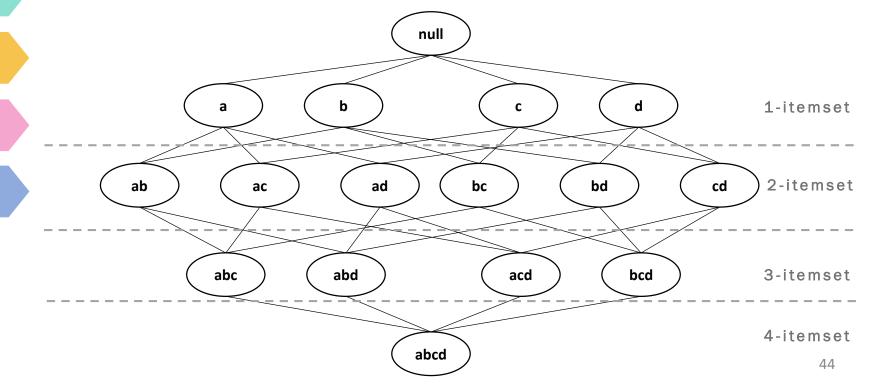


### Mining correlations





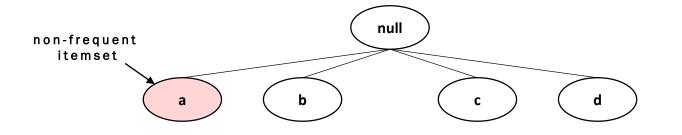
### Mining correlations





### Mining correlations

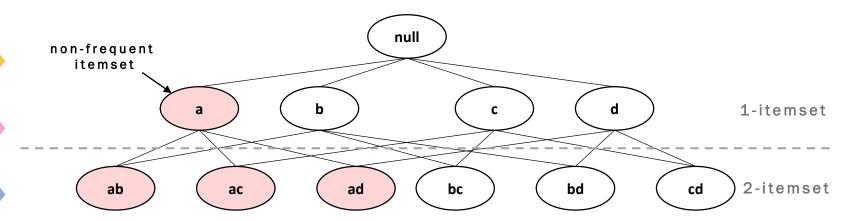
Minimum Thresholds must be defined to avoid unnecessary processing of non-frequent itemsets:



1-itemset

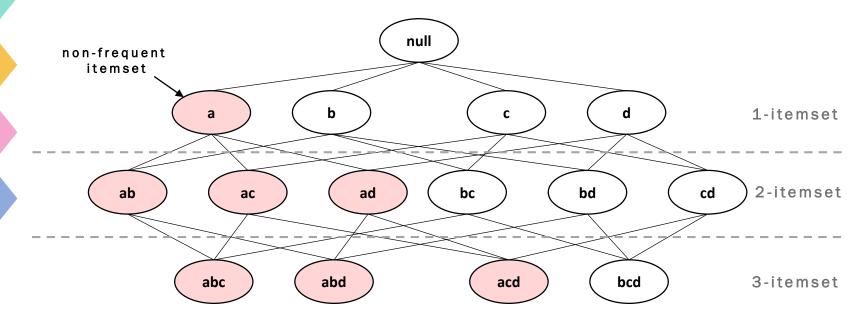


### Mining correlations



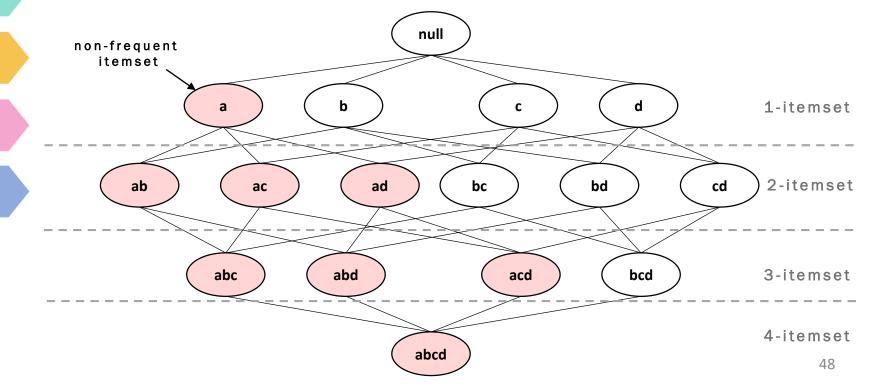


### Mining correlations



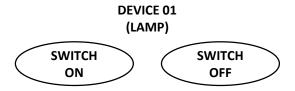


#### Mining correlations





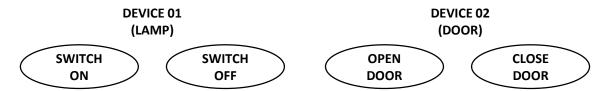
#### Mining correlations





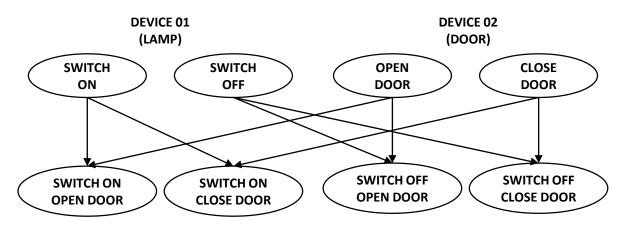


#### Mining correlations





### Mining correlations





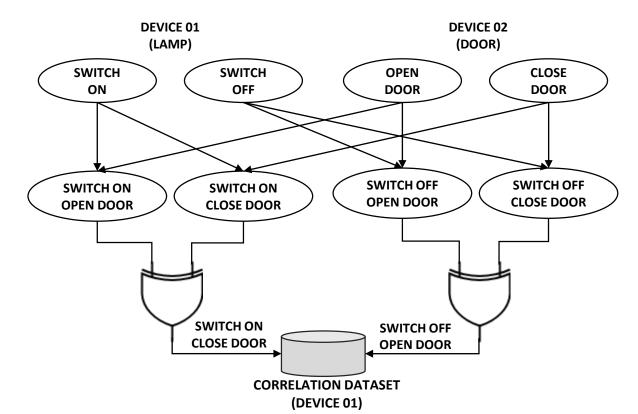








#### Mining correlations

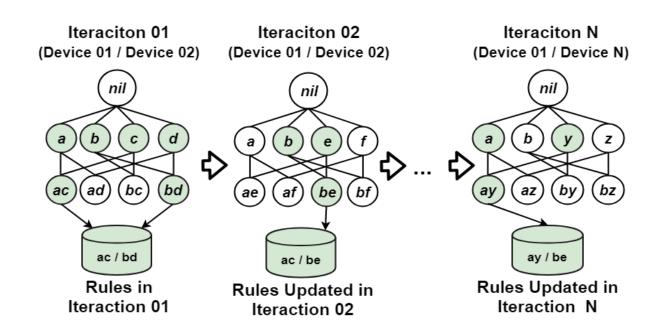






#### Mining correlations

This process repeats to each device in network updating the most relevant rules in correlation dataset



53



#### What should I do now!?

Based on those rules it is possible to Device 01 perform a request

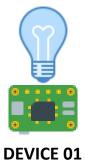
to Device 02 change its states to satisfy the rule;

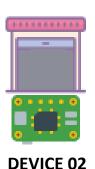




#### What should I do now!?

Based on those rules it is possible to Device O1 perform a request to Device O2 change its states to satisfy the rule;









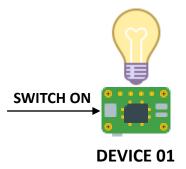


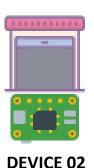




#### What should I do now!?

Based on those rules it is possible to Device O1 perform a request to Device O2 change its states to satisfy the rule;









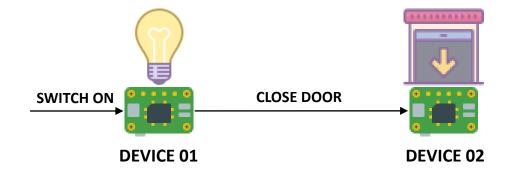






#### What should I do now!?

Based on those rules it is possible to Device O1 perform a request to Device O2 change its states to satisfy the rule;







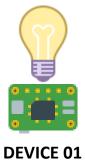


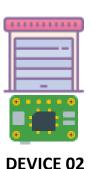




#### What should I do now!?

Based on those rules it is possible to Device 01 perform a request to Device 02 change its states to satisfy the rule;







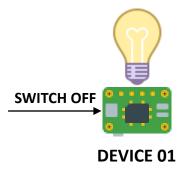


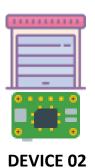




#### What should I do now!?

Based on those rules it is possible to Device O1 perform a request to Device O2 change its states to satisfy the rule;











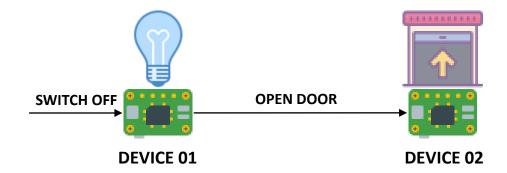






#### What should I do now!?

Based on those rules it is possible to Device O1 perform a request to Device O2 change its states to satisfy the rule;









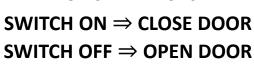


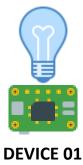


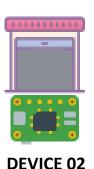
#### What should I do now!?

Based on those rules it is possible to Device 01 perform a request to Device 02 change its states to satisfy the rule;

**RULES IN DEVICE 01** 



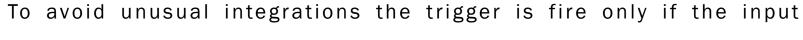




**ALWAYS!?** THAT'S SOUND STUPID DANGEROUS



#### Slot restrictions



action were the most probable to occur at the current slot.



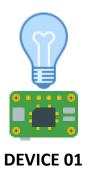


#### Slot restrictions

To avoid unusual integrations the trigger is fire only if the input action were the most probable to occur at the current slot.

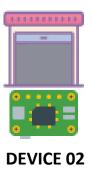
## **RULES IN DEVICE 01**

SWITCH ON ⇒ CLOSE DOOR
SWITCH OFF $\Rightarrow$ OPEN DOOF



00:00	SWITCH OFF
01:00	-
02:00	-
21:00	SWITCH ON
22:00	SWITCH ON
23:00	SWITCH OFF

00:00	CLOSE DOOR
01:00	CLOSE DOOR
02:00	-
•••	
21:00	OPEN DOOR
21:00	OPEN DOOR CLOSE DOOR

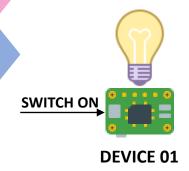




#### Slot restrictions

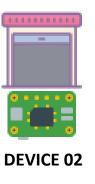
To avoid unusual integrations the trigger is fire only if the input action were the most probable to occur at the current slot.

RULES IN DEVICE 01
SWITCH ON $\Rightarrow$ CLOSE DOOR
SWITCH OFF ⇒ OPEN DOOR



00:00	SWITCH OFF
01:00	-
02:00	-
21:00	SWITCH ON
22:00	SWITCH ON
23:00	SWITCH OFF

00:00	CLOSE DOOR
01:00	CLOSE DOOR
02:00	-
21:00	OPEN DOOR
21:00 22:00	OPEN DOOR CLOSE DOOR



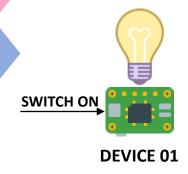


#### Slot restrictions

To avoid unusual integrations the trigger is fire only if the input action were the most probable to occur at the current slot.

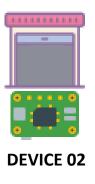
## **RULES IN DEVICE 01**

WITCH ON ⇒ CLOSE DOOR
WITCH OFF ⇒ OPEN DOOR



00:00	SWITCH OFF	×
01:00	-	
02:00	-	
21:00	SWITCH ON	
22:00	SWITCH ON	
23:00	SWITCH OFF	X

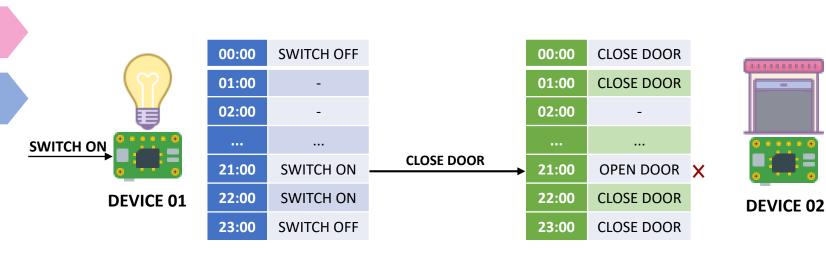
00:00	CLOSE DOOR
01:00	CLOSE DOOR
02:00	-
•••	
21:00	OPEN DOOR
21:00	OPEN DOOR CLOSE DOOR





#### Slot restrictions

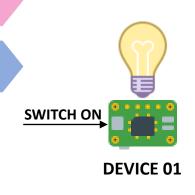
To avoid unusual integrations the trigger is fire only if the input action were the most probable to occur at the current slot.



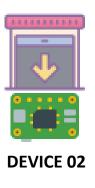


#### Slot restrictions

To avoid unusual integrations the trigger is fire only if the input action were the most probable to occur at the current slot.



00:00	SWITCH OFF		00:00	CLOSE DOOR
01:00	-		01:00	CLOSE DOOR
02:00	-		02:00	-
21:00	SWITCH ON		21:00	OPEN DOOR
22:00	SWITCH ON	CLOSE DOOR	22:00	CLOSE DOOR
23:00	SWITCH OFF		23:00	CLOSE DOOR





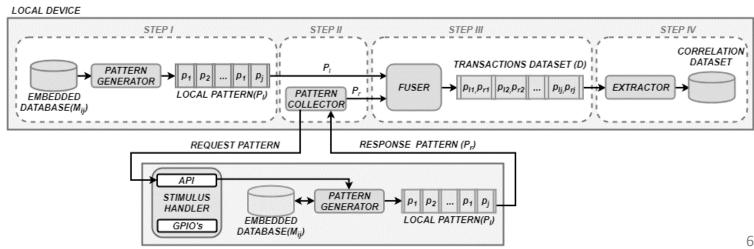
#### **Process Overview**

Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions.











CORRELATION DATASET

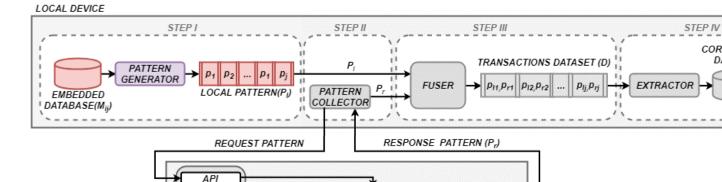
69

#### **Process Overview**

Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions.

Step I: Identify the local pattern





**PATTERN** 

GENERATOR

p<sub>1</sub> p<sub>2</sub> ... p<sub>1</sub> p<sub>j</sub>

LOCAL PATTERN(P<sub>i</sub>)

(NEXT) REMOTE DEVICE FROM TARGET'S LIST

EMBEDDED DATABASE(M<sub>ii</sub>)

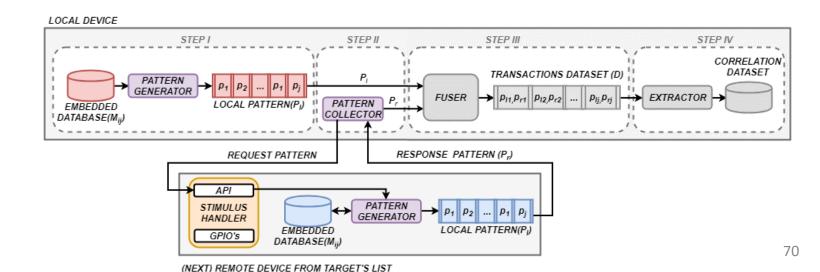
STIMULUS

**HANDLER** 



#### **Process Overview**

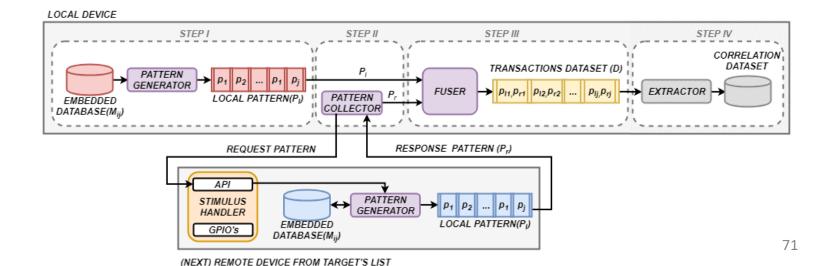
- Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions.
- Step I: Identify the local pattern
- Step II: Obtain the (next)remote pattern (http request/response)





#### **Process Overview**

- Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions.
- Step I: Identify the local pattern
- Step II: Obtain the (next)remote pattern (http request/response)
- Step III: Fusion the patterns (create a transaction dataset)



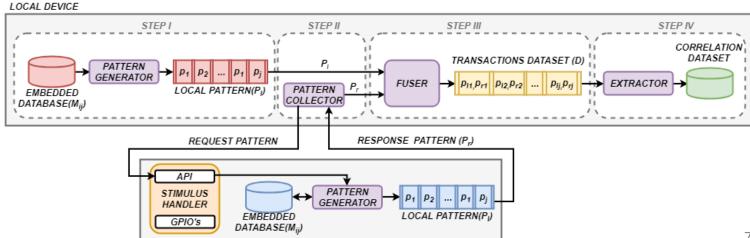


#### **Process Overview**

- Once that each devices satisfies the previously specifications is possible to identify implicit correlations between its actions..
- Step I: Identify the local pattern
- Step II: Obtain the (next)remote pattern (http request/response)
- Step III: Fusion the patterns (create a transaction dataset)

(NEXT) REMOTE DEVICE FROM TARGET'S LIST

Step IV: Extract the correlations (Associative rule mining)(\*Back to II)



72

### **EXPERIMENTS**



#### **Experiment**

Identify correlations between two devices.

Teste the integration between their actions (if exists).

Implemented in Lua (https://nodemcu-build.com/)







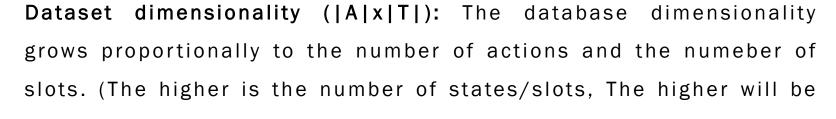






#### **Important Notes**

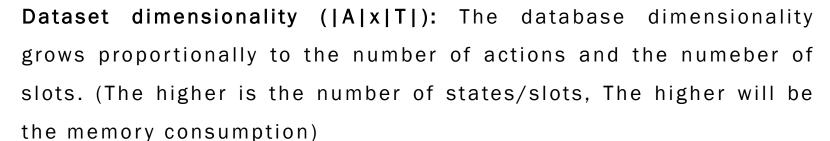
the memory consumption)







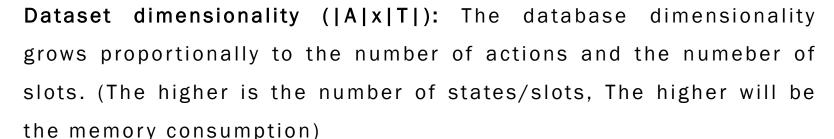
#### **Important Notes**



**Pattern sharing:** All devices must be abble to handle its own pattens and dataset AND a remote device's pattern.(patterns fusion and associative analysis)



#### Important Notes

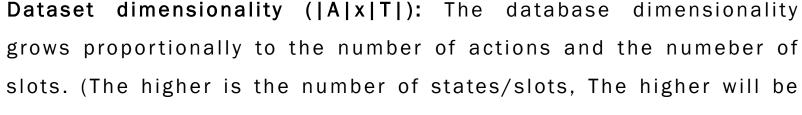


**Pattern sharing:** All devices must be abble to handle its own pattens and dataset AND a remote device's pattern.(patterns fusion and associative analysis)

**Slots Interval:** The slot interval must be the same for all devices. That implies in equal |T| for all devices.



#### Important Notes



the memory consumption)

**Pattern sharing:** All devices must be abble to handle its own pattens and dataset AND a remote device's pattern.(patterns fusion and associative analysis)

**Slots Interval:** The slot interval must be the same for all devices. That implies in equal |T| for all devices.

Device discovery: Multicast uses UDP!



#### Federal University of Amazonas

Institute of Computing Graduate Program in Informatics Interest Group on Embedded Systems

#### **QUESTIONS?**

Manaus, July 13<sup>th</sup> 2018



#### Federal University of Amazonas

Institute of Computing Graduate Program in Informatics Interest Group on Embedded Systems

#### **THANK YOU**

Manaus, July 13<sup>th</sup> 2018

### TRIGGERS CORRELATIONS



#### **Associative Analysis Metrics**

Support: Number of elements in a set

#### **CENTRALIZED**

PAO	LEITE	CAFÉ
1	1	1
1	1	1
1	0	1
0	1	1
1	1	0
0	1	0

(BREAD->MILK): 3/6 (BREAD->COFFEE): 3/6 (MILK->COFFEE): 3/6

PAO	LEITE
1	1
1	1
1	0
0	1
1	1
0	1

(BREAD->MILK) 3/6

#### **DECENTRALIZED**

	PAO	CAFÉ	
	1	1	
	1	1	
	1	1	
	0	1	
	1	0	
BREAD->COFFEE			

(B 3/5

LEITE	CAFÉ
1	1
1	1
0	1
1	1
1	0
1	0

(BREAD->COFFEE) 3/6

### TRIGGERS CORRELATIONS



### **Associative Analysis Metrics**

- Support: Number of elements in a set
- Define Max Transaction Value ( |T| = 7 )

#### **CENTRALIZED**

PAO	LEITE	CAFÉ
1	1	1
1	1	1
1	0	1
0	1	1
1	1	0
0	1	0

(BREAD->MILK): 2/6 (BREAD->COFFEE): 2/6 (MILK->COFFEE): 3/6

## PAO

1

	_
1	1
1	0
0	1
1	1
0	1

LEITE

1

(BREAD->MILK) 2/7

#### **DECENTRALIZED**

PAO	CAFÉ
1	1
1	1
1	1
0	1
1	0
-	-
-	-

LEITE	CAFÉ
1	1
1	1
0	1
1	1
1	0
1	0
-	-

(BREAD->COFFEE) (BREAD->COFFEE) 2/7 3/7