Insight in Cyber Safety when Remotely Operating SCADA Systems of Dutch Critical Infrastructure Objects

Presented by: Tina Tami
Supervisor: Cedric Both

NEXT LEVEL SPECIALISTS
DATADIGEST

Introduction

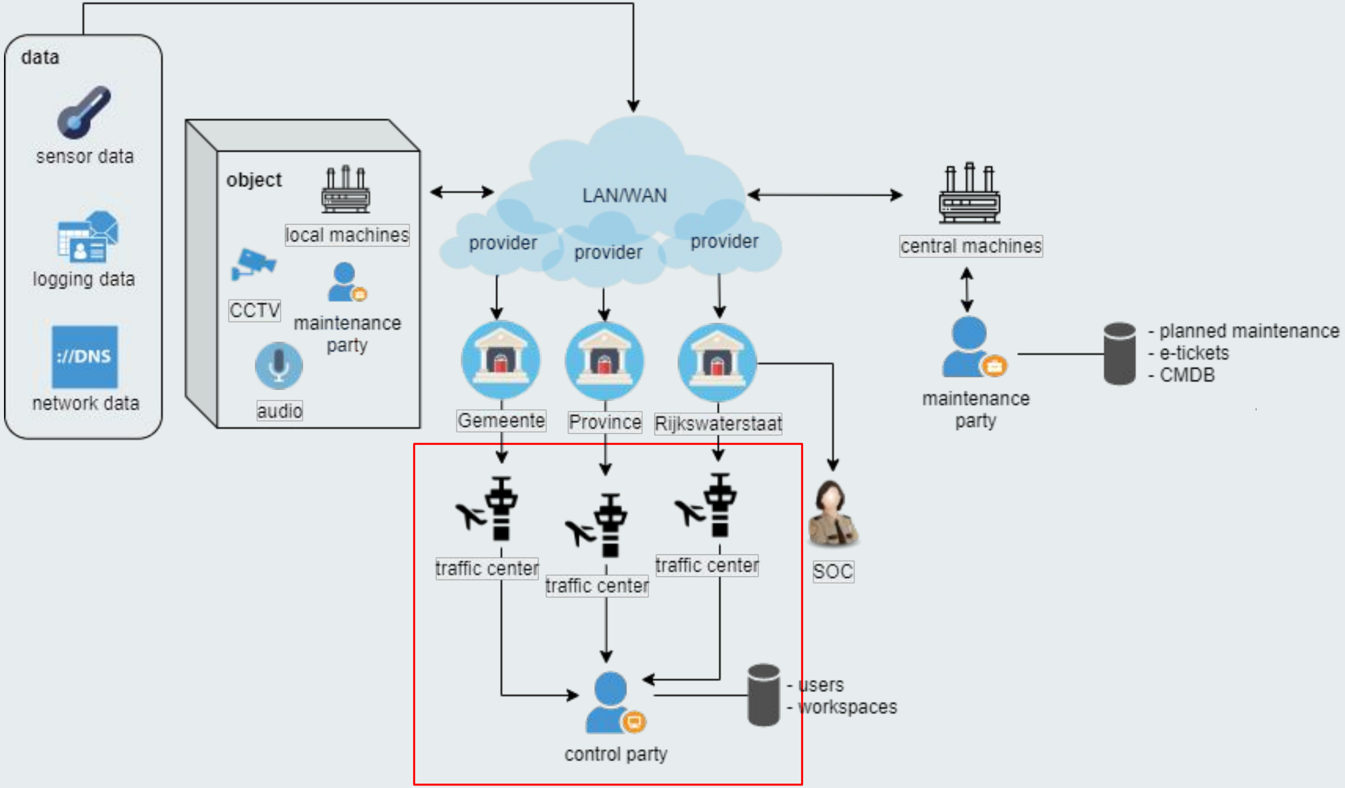
Supervisory **C**ontrol **a**nd **D**ata **A**cquisition: the collection, forwarding, processing and visualization of measurement and control signals from different machines in large industrial systems

→ Bridges

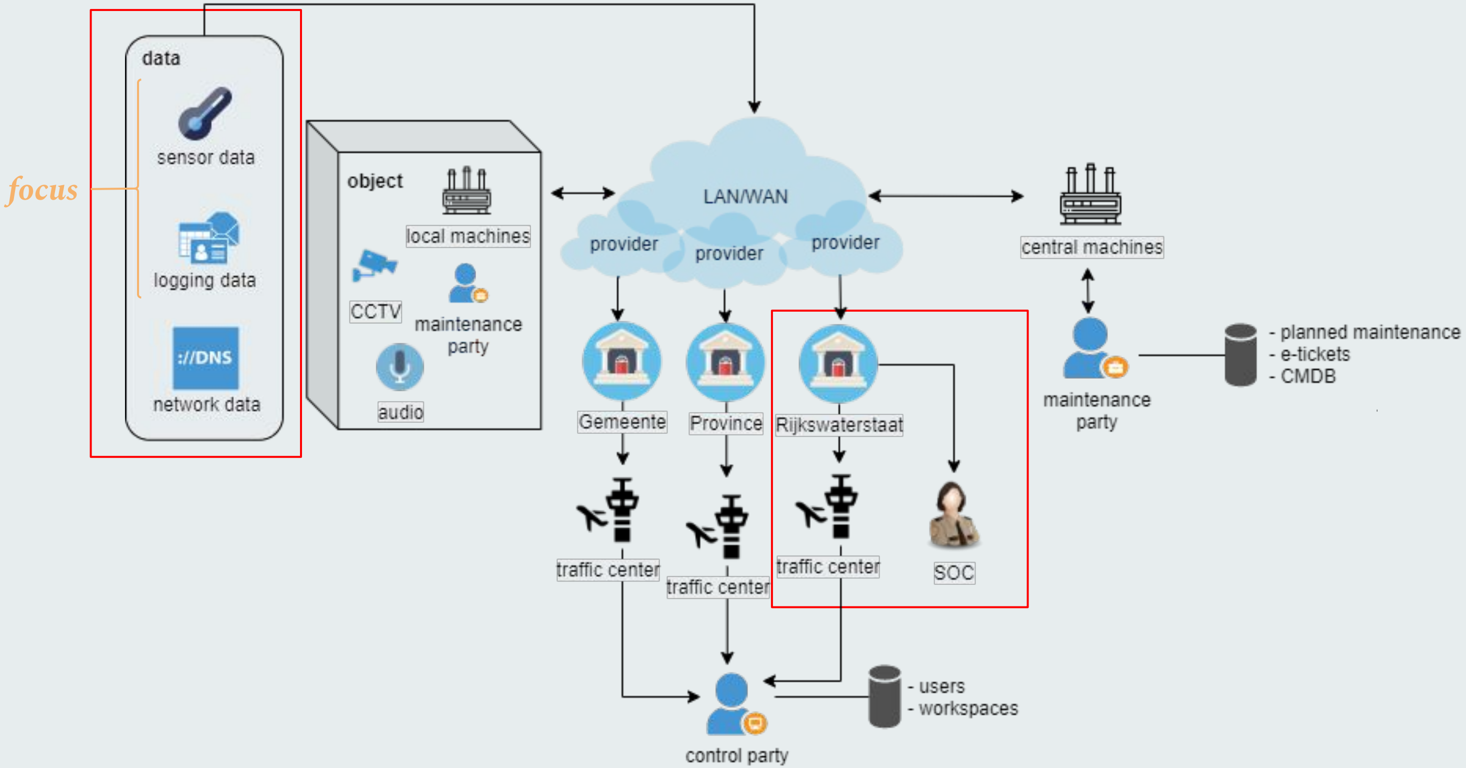
→ Tunnels

→ Locks

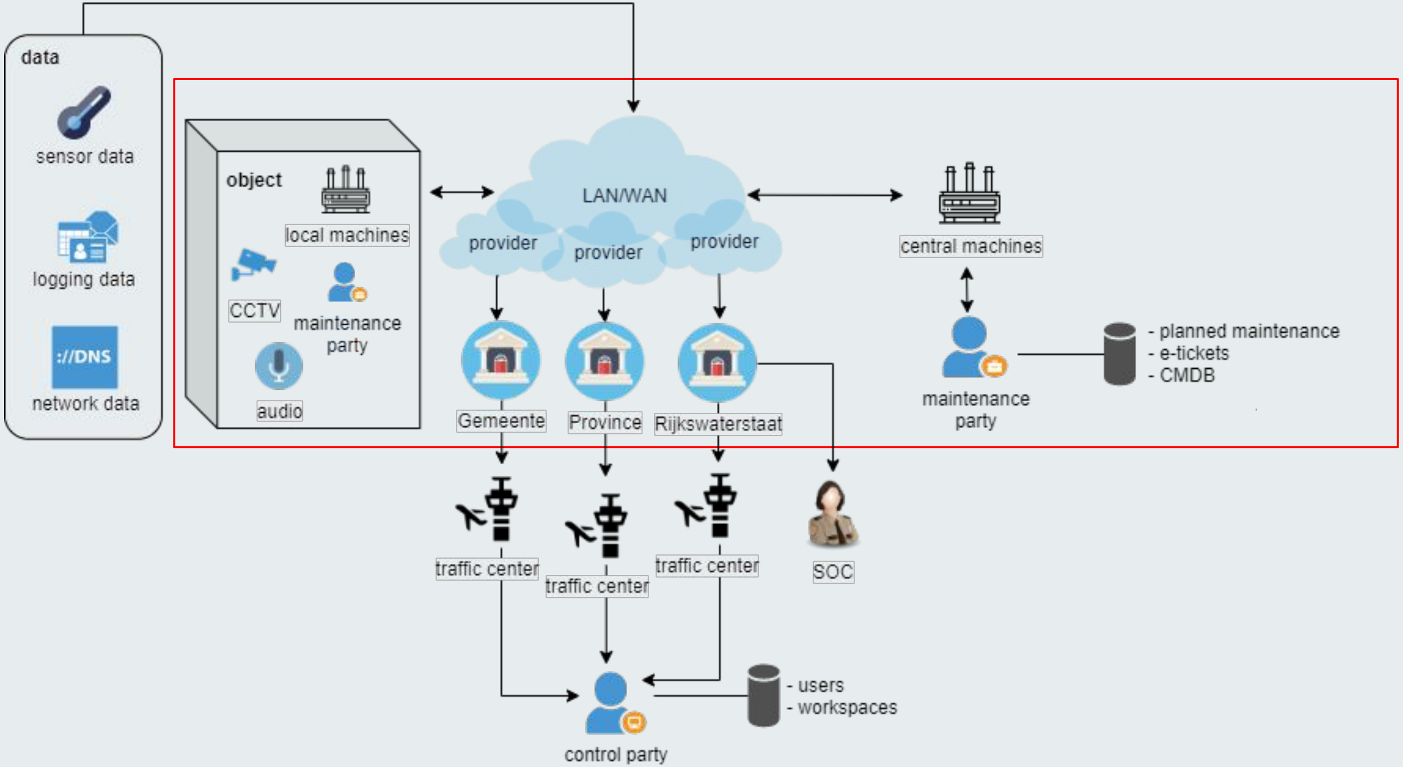
Introduction



Introduction



Introduction



Research questions

How can anomalies be detected in event data from SCADA and other involved systems, in order to provide security alerting and decision support rules?

- What data is relevant and necessary in order to decide if an action is considered anomalous?
- What techniques can be used to detect anomalies, and which one would work best in this case?

Related work

- Use NARX (nonlinear autoregressive exogenous model) to estimate temperature signals of wind turbines - Y. Cui et al.
- Anomaly detection in SCADA systems using flow whitelisting - R. Ramos et al.

Methodology

Necessary data

- Performed actions → legal?
- Time and location → strange?
- Agent → qualified?

Methodology

Necessary data

- Performed actions → legal?
- Time and location → strange?
- Agent → qualified?

Approach requirements

- Handle textual input
- Event-based data
- Take into account previous event(s)
- Real-time application

Anomaly Detection Algorithms

SVM using TF-IDF

- Support-Vector Machine: Find natural clustering of the data to groups

- Term Frequency–Inverse Document Frequency: Reflect how important a word is to a document in a corpus

🚫 Not what we are looking for

Clustering algorithms

- K-means clustering
- Gaussian mixture model

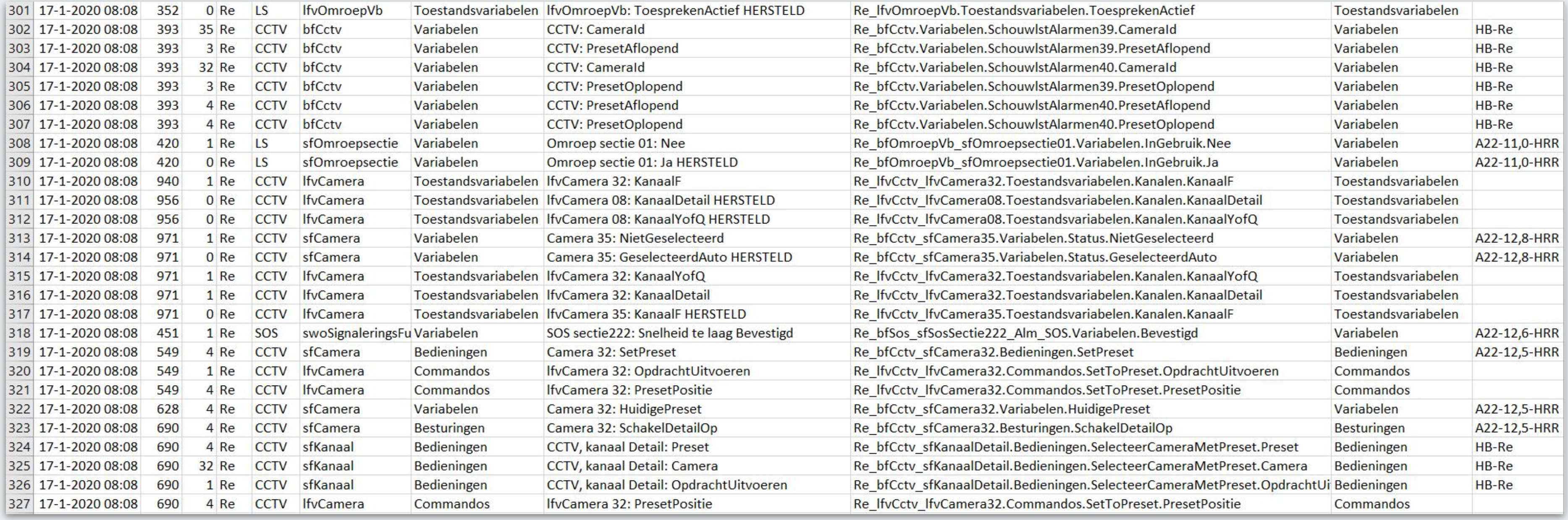
🚫 Numeric data

Markov Chain Model

- Probability of each observation depends on the state attained in the previous observation

➤ Interesting?

Data set



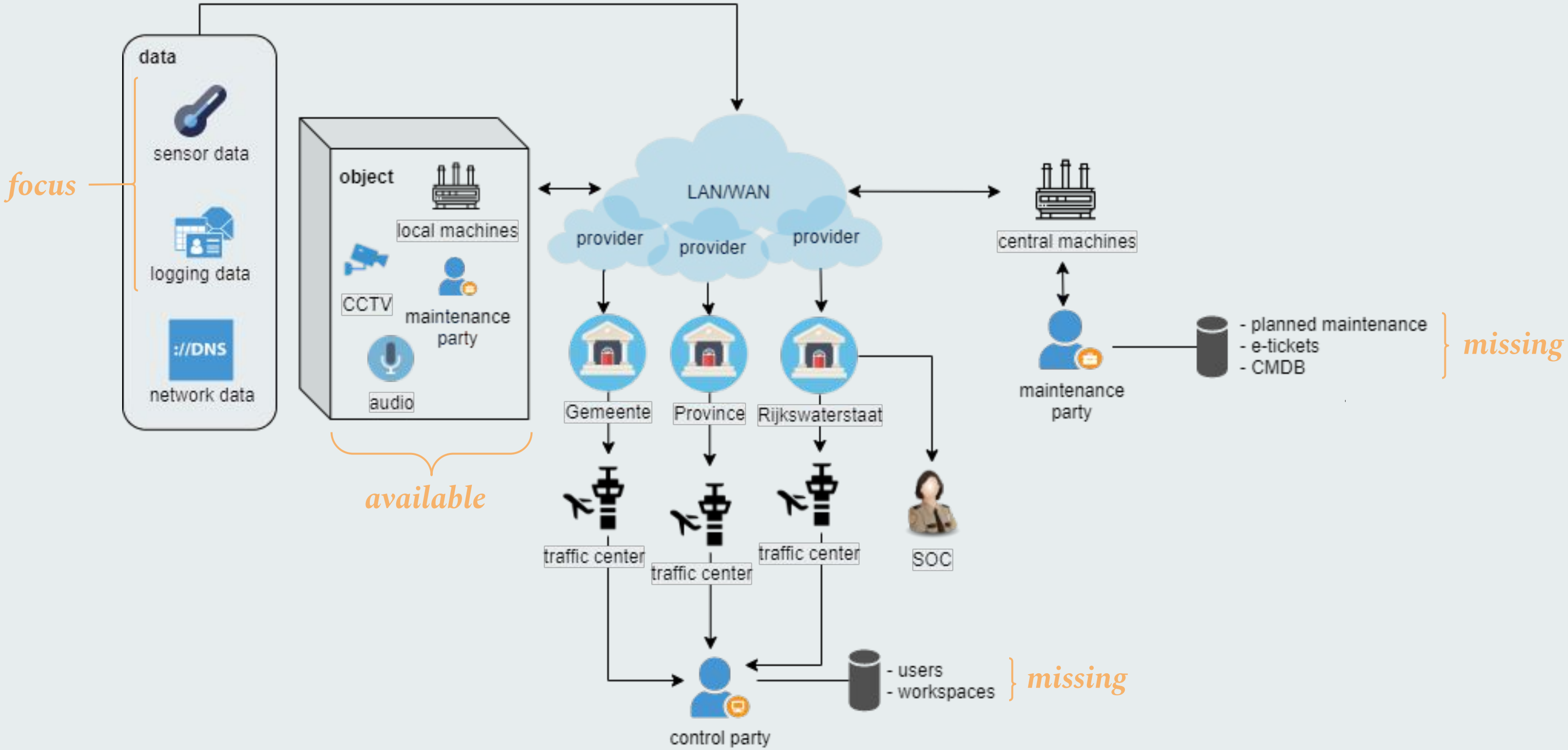
301	17-1-2020 08:08	352	0	Re	LS	IfvOmroepVb	Toestandsvariabelen	IfvOmroepVb: ToesprekenActief HERSTELD	Re_IfvOmroepVb.Toestandsvariabelen.ToesprekenActief	Toestandsvariabelen	
302	17-1-2020 08:08	393	35	Re	CCTV	bfCctv	Variabelen	CCTV: CameraId	Re_bfCctv.Variabelen.SchouwlstAlarmen39.CameraId	Variabelen	HB-Re
303	17-1-2020 08:08	393	3	Re	CCTV	bfCctv	Variabelen	CCTV: PresetAflopend	Re_bfCctv.Variabelen.SchouwlstAlarmen39.PresetAflopend	Variabelen	HB-Re
304	17-1-2020 08:08	393	32	Re	CCTV	bfCctv	Variabelen	CCTV: CameraId	Re_bfCctv.Variabelen.SchouwlstAlarmen40.CameraId	Variabelen	HB-Re
305	17-1-2020 08:08	393	3	Re	CCTV	bfCctv	Variabelen	CCTV: PresetOplopend	Re_bfCctv.Variabelen.SchouwlstAlarmen39.PresetOplopend	Variabelen	HB-Re
306	17-1-2020 08:08	393	4	Re	CCTV	bfCctv	Variabelen	CCTV: PresetAflopend	Re_bfCctv.Variabelen.SchouwlstAlarmen40.PresetAflopend	Variabelen	HB-Re
307	17-1-2020 08:08	393	4	Re	CCTV	bfCctv	Variabelen	CCTV: PresetOplopend	Re_bfCctv.Variabelen.SchouwlstAlarmen40.PresetOplopend	Variabelen	HB-Re
308	17-1-2020 08:08	420	1	Re	LS	sfOmroepsectie	Variabelen	Omroep sectie 01: Nee	Re_bfOmroepVb_sfOmroepsectie01.Variabelen.InGebruik.Nee	Variabelen	A22-11,0-HRR
309	17-1-2020 08:08	420	0	Re	LS	sfOmroepsectie	Variabelen	Omroep sectie 01: Ja HERSTELD	Re_bfOmroepVb_sfOmroepsectie01.Variabelen.InGebruik.Ja	Variabelen	A22-11,0-HRR
310	17-1-2020 08:08	940	1	Re	CCTV	IfvCamera	Toestandsvariabelen	IfvCamera 32: KanaalF	Re_IfvCctv_IfvCamera32.Toestandsvariabelen.Kanalen.KanaalF	Toestandsvariabelen	
311	17-1-2020 08:08	956	0	Re	CCTV	IfvCamera	Toestandsvariabelen	IfvCamera 08: KanaalDetail HERSTELD	Re_IfvCctv_IfvCamera08.Toestandsvariabelen.Kanalen.KanaalDetail	Toestandsvariabelen	
312	17-1-2020 08:08	956	0	Re	CCTV	IfvCamera	Toestandsvariabelen	IfvCamera 08: KanaalYofQ HERSTELD	Re_IfvCctv_IfvCamera08.Toestandsvariabelen.Kanalen.KanaalYofQ	Toestandsvariabelen	
313	17-1-2020 08:08	971	1	Re	CCTV	sfCamera	Variabelen	Camera 35: NietGeselecteerd	Re_bfCctv_sfCamera35.Variabelen.Status.NietGeselecteerd	Variabelen	A22-12,8-HRR
314	17-1-2020 08:08	971	0	Re	CCTV	sfCamera	Variabelen	Camera 35: GeselecteerdAuto HERSTELD	Re_bfCctv_sfCamera35.Variabelen.Status.GeselecteerdAuto	Variabelen	A22-12,8-HRR
315	17-1-2020 08:08	971	1	Re	CCTV	IfvCamera	Toestandsvariabelen	IfvCamera 32: KanaalYofQ	Re_IfvCctv_IfvCamera32.Toestandsvariabelen.Kanalen.KanaalYofQ	Toestandsvariabelen	
316	17-1-2020 08:08	971	1	Re	CCTV	IfvCamera	Toestandsvariabelen	IfvCamera 32: KanaalDetail	Re_IfvCctv_IfvCamera32.Toestandsvariabelen.Kanalen.KanaalDetail	Toestandsvariabelen	
317	17-1-2020 08:08	971	0	Re	CCTV	IfvCamera	Toestandsvariabelen	IfvCamera 35: KanaalF HERSTELD	Re_IfvCctv_IfvCamera35.Toestandsvariabelen.Kanalen.KanaalF	Toestandsvariabelen	
318	17-1-2020 08:08	451	1	Re	SOS	swoSignaleringsFu	Variabelen	SOS sectie222: Snelheid te laag Bevestigd	Re_bfSos_sfSosSectie222_Alm_SOS.Variabelen.Bevestigd	Variabelen	A22-12,6-HRR
319	17-1-2020 08:08	549	4	Re	CCTV	sfCamera	Bedieningen	Camera 32: SetPreset	Re_bfCctv_sfCamera32.Bedieningen.SetPreset	Bedieningen	A22-12,5-HRR
320	17-1-2020 08:08	549	1	Re	CCTV	IfvCamera	Commandos	IfvCamera 32: OpdrachtUitvoeren	Re_IfvCctv_IfvCamera32.Commandos.SetToPreset.OpdrachtUitvoeren	Commandos	
321	17-1-2020 08:08	549	4	Re	CCTV	IfvCamera	Commandos	IfvCamera 32: PresetPositie	Re_IfvCctv_IfvCamera32.Commandos.SetToPreset.PresetPositie	Commandos	
322	17-1-2020 08:08	628	4	Re	CCTV	sfCamera	Variabelen	Camera 32: HuidigePreset	Re_bfCctv_sfCamera32.Variabelen.HuidigePreset	Variabelen	A22-12,5-HRR
323	17-1-2020 08:08	690	4	Re	CCTV	sfCamera	Besturingen	Camera 32: SchakelDetailOp	Re_bfCctv_sfCamera32.Besturingen.SchakelDetailOp	Besturingen	A22-12,5-HRR
324	17-1-2020 08:08	690	4	Re	CCTV	sfKanaal	Bedieningen	CCTV, kanaal Detail: Preset	Re_bfCctv_sfKanaalDetail.Bedieningen.SelecteerCameraMetPreset.Preset	Bedieningen	HB-Re
325	17-1-2020 08:08	690	32	Re	CCTV	sfKanaal	Bedieningen	CCTV, kanaal Detail: Camera	Re_bfCctv_sfKanaalDetail.Bedieningen.SelecteerCameraMetPreset.Camera	Bedieningen	HB-Re
326	17-1-2020 08:08	690	1	Re	CCTV	sfKanaal	Bedieningen	CCTV, kanaal Detail: OpdrachtUitvoeren	Re_bfCctv_sfKanaalDetail.Bedieningen.SelecteerCameraMetPreset.OpdrachtUi	Bedieningen	HB-Re
327	17-1-2020 08:08	690	4	Re	CCTV	IfvCamera	Commandos	IfvCamera 32: PresetPositie	Re_IfvCctv_IfvCamera32.Commandos.SetToPreset.PresetPositie	Commandos	

Data set

Event logging

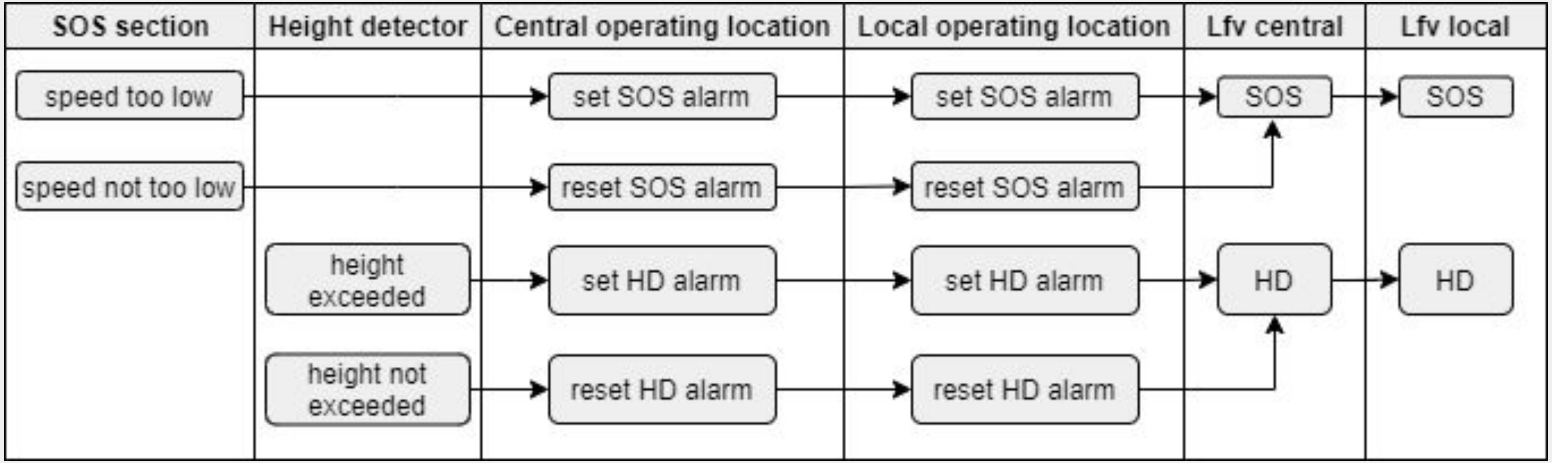
- *24* unique subsystems → *3* unique subsystems
- *2191* unique text entries → *23* unique text entries

timestamp	**subsystem**	**text**	**location**
17-1-2020 08:00:26:784	HD	Hoogte detector 03: HoogteOverschrijding Nee HERSTELD	A208b-10,7


data
sensor data
logging data
network data
://DNS
focus
object
local machines
CCTV
maintenance party
audio
available
LAN/WAN
provider
provider
provider
Gemeente
Province
Rijkswaterstaat
traffic center
traffic center
traffic center
SOC
control party
- users
- workspaces
missing
central machines
maintenance party
- planned maintenance
- e-tickets
- CMDB
missing

Data set

Event logging



Discrete-Time Markov Chain

Sequence of observations

$\{X_1, X_2, \dots, X_n\}$

such that

$P_{ij} = P(X_{t+1} = j | X_t = i)$

$P_{ij} = n_{ij}/n_i$

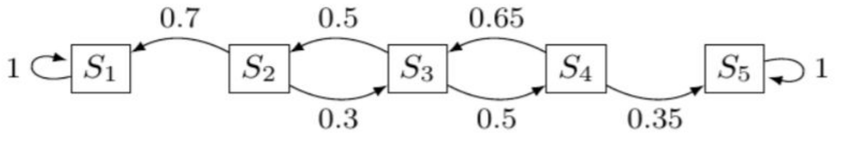
Transition probability matrix

$$\mathbf{P} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1s} \\ P_{21} & P_{22} & \cdots & P_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s1} & P_{s2} & \cdots & P_{ss} \end{bmatrix}$$

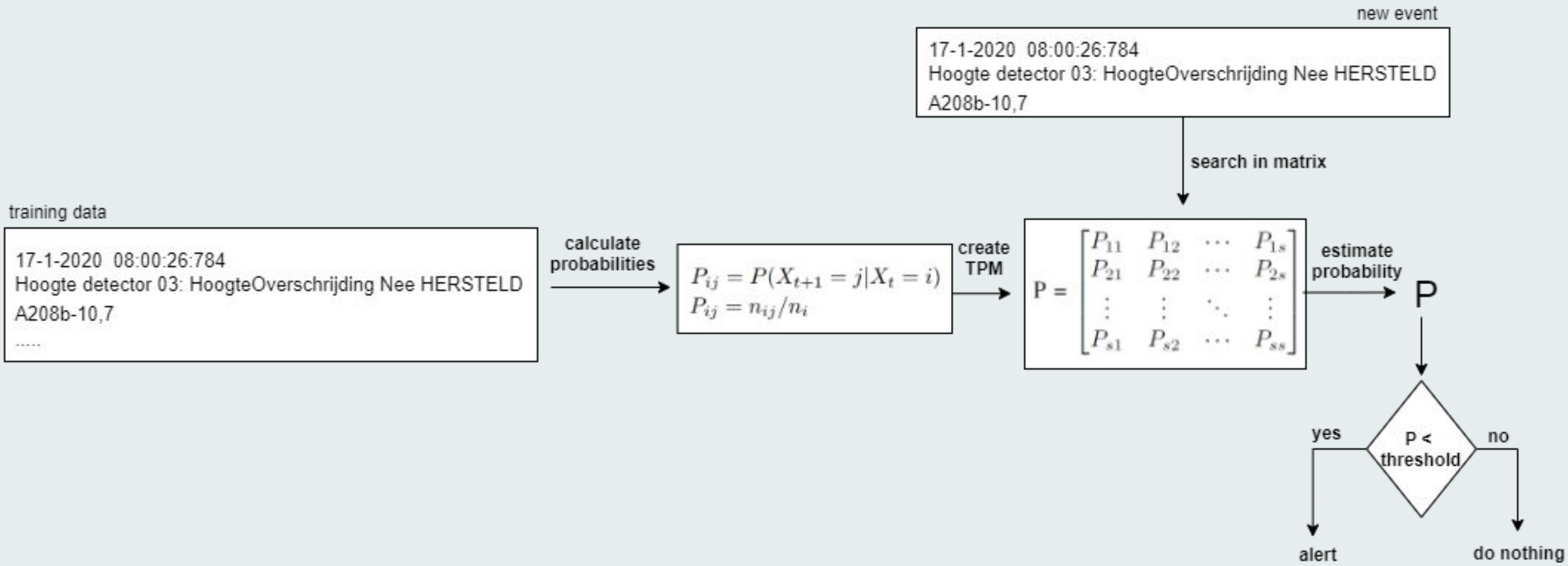
- The probability of transitioning to any particular state is dependent solely on the current state

- Estimate the probability of new events based on the transition probability matrix

Diagram



Approach



training data

```
17-1-2020  08:00:26:784
Hoogte detector 03: HoogteOverschrijding Nee HERSTELD
A208b-10,7
.....
```

calculate probabilities

$$P_{ij} = P(X_{t+1} = j | X_t = i)$$
$$P_{ij} = n_{ij}/n_i$$

create TPM

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1s} \\ P_{21} & P_{22} & \cdots & P_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ P_{s1} & P_{s2} & \cdots & P_{ss} \end{bmatrix}$$

new event

```
17-1-2020  08:00:26:784
Hoogte detector 03: HoogteOverschrijding Nee HERSTELD
A208b-10,7
```

search in matrix

estimate probability

P

P < threshold

yes

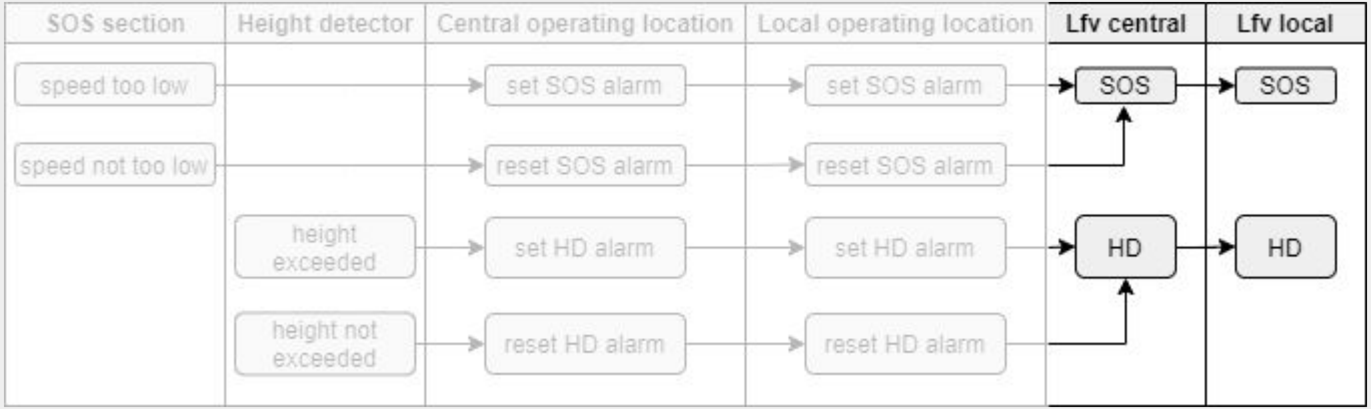
alert

no

do nothing

Approach

Definite anomalies



SOS section

Height detector

Central operating location

Local operating location

Lfv central

Lfv local

speed too low

set SOS alarm

set SOS alarm

SOS

SOS

speed not too low

reset SOS alarm

reset SOS alarm

height exceeded

set HD alarm

set HD alarm

HD

HD

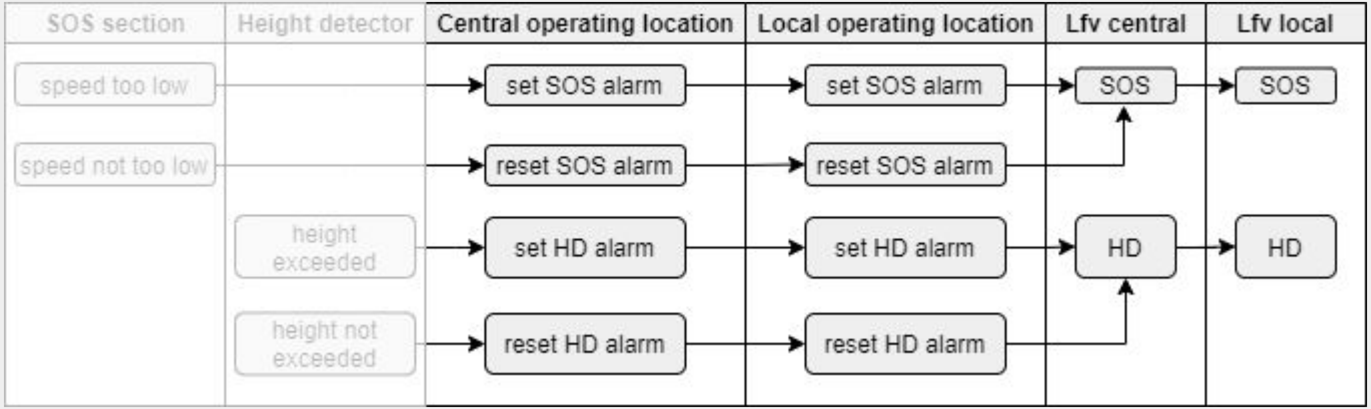
height not exceeded

reset HD alarm

reset HD alarm

Approach

Possible anomalies



SOS section | Height detector | Central operating location | Local operating location | Lfv central | Lfv local

speed too low → set SOS alarm → set SOS alarm → SOS → SOS

speed not too low → reset SOS alarm → reset SOS alarm → SOS

height exceeded → set HD alarm → set HD alarm → HD → HD

height not exceeded → reset HD alarm → reset HD alarm → HD

Results

Output: test set

```
-
```

Output: test set + anomalies

```
#1
Current state:    Snelheid te laag
Next state:  lfvBediening Centraal: SOS
Probability of this happening:  0.0
Time of anomaly:  17-1-2020 17:53:39:788
Location of anomaly:  A22-12,4-HRL
-----
#2
Current state:  Bedienlocatie Lokaal: SetAlarmContact SOS
Next state:    Snelheid te laag HERSTELD
Probability of this happening:  0.0
Time of anomaly:  17-1-2020 18:07:59:998
Location of anomaly:  TN
```

Results

Output: test set

```
#1
Current state:  lfvBediening Centraal: SOS
Next state:   HoogteOverschrijding Ja
Probability of this happening:  0.021505376344086023
Time of anomaly:  17-1-2020 18:27:56:932
Location of anomaly:  nan
-----
#2
Current state:  lfvBediening Centraal: SOS
Next state:   HoogteOverschrijding Ja
Probability of this happening:  0.021505376344086023
Time of anomaly:  17-1-2020 10:45:40:456
Location of anomaly:  nan
```

Output: test set + anomalies

```
#3
Current state:  lfvBediening Centraal: SOS
Next state:  Bedienlocatie Centraal: SetAlarmContact
DefHoogteDetectie
Probability of this happening:  0.010752688172043012
Time of anomaly:  17-1-2020 18:27:56:932
Location of anomaly:  nan
-----
#4
Current state:  lfvBediening Centraal: SOS
Next state:   HoogteOverschrijding Ja
Probability of this happening:  0.021505376344086023
Time of anomaly:  17-1-2020 10:45:40:456
Location of anomaly:  nan
```

Results

Output: test set

```
#3
Current state:  lfvBediening Lokaal: SOS
Next state:  Bedienlocatie Centraal:
ResetAlarmContact SOS
Probability of this happening:  0.06989247311827956
Time of anomaly:  17-1-2020 17:53:39:854
Location of anomaly:  nan
-----
#4
Current state:  lfvBediening Lokaal: SOS
Next state:  Bedienlocatie Centraal:
ResetAlarmContact SOS
Probability of this happening:  0.06989247311827956
Time of anomaly:  17-1-2020 18:12:17:34
Location of anomaly:  nan
```

Output: test set + anomalies

```
#5
Current state:    Snelheid te laag
Next state:  Bedienlocatie Centraal: ResetAlarmContact SOS
Probability of this happening:  0.02608695652173913
Time of anomaly:  17-1-2020 17:42:03:781
Location of anomaly:  A22-11,4-HRR
-----
#6
Current state:  lfvBediening Lokaal: DefHoogteDetectie
Next state:  Bedienlocatie Centraal: ResetAlarmContact
DefHoogteDetectie
Probability of this happening:  0.045454545454545456
Time of anomaly:  17-1-2020 19:32:38:961
Location of anomaly:  nan
```

Discussion

- Real-time application
- Difficult to evaluate
- Data used for transition probability matrix has to be **complete** and **reliable**
- Which threshold?
- Relevant data is missing

Conclusion

How can anomalies be detected in event data from SCADA and other involved systems, in order to provide security alerting and decision support rules?

- What data is relevant and necessary in order to decide if an action is considered anomalous?
  → **Event that is happening, time, location, user, schedule**
  → **Missing: agent, schedule**
- What techniques can be used to detect anomalies, and which one would work best in this case?
  → **Markov Chain**

Future work

- Add decision support by taking into account extra data
- Method for finding ideal threshold