Synthesizing Stealthy Reprogramming Attacks on Cardiac Devices

to appear in IEEE/ACM International Conference Cyber-Physical Systems (ICCPS 2019)

Nicola Paoletti

Royal Holloway, University of London

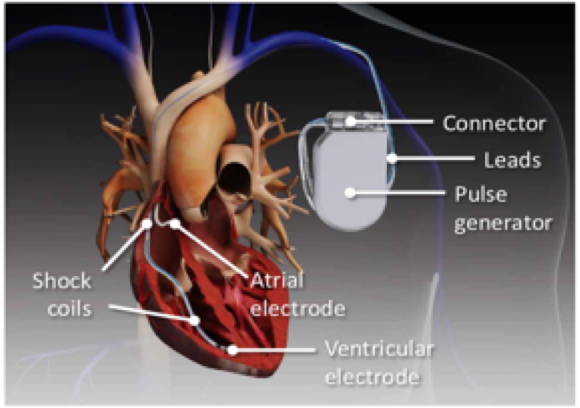
Joint work with:

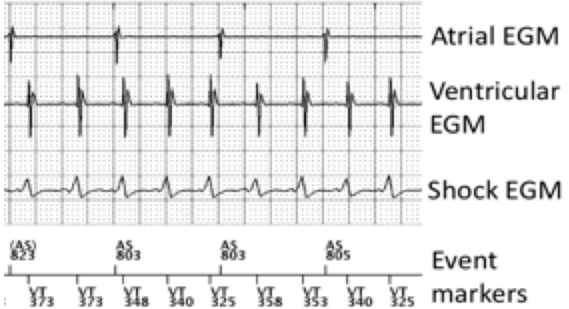
Scott A Smolka, Shan Lin, Zachary Gruber (Stony Brook), Zhihao Jiang (ShangaiTech),
Md Ariful Islam (Texas Tech), Rahul Mangharam, Houssam Abbas (UPenn)

ISG Research Seminar, RHUL, 28 March 2019

What are ICDs?

- Implantable cardioverter defibrillator
  - Prevent sudden cardiac death in patients
  - Pacemaker and defibrillator function
- ICD therapy
  - Monitor 3 signals: atrial, ventricular, shock EGM
  - ATP – Anti-tachycardia pacing
  - **High-energy shocks**

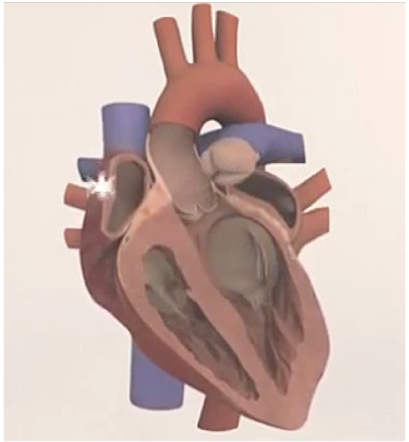




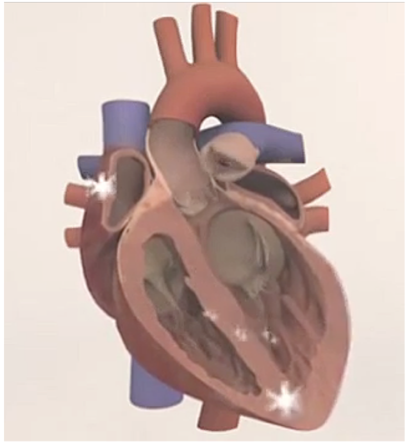
What are ICDs?

ICDs execute **discrimination algorithms** to distinguish between:

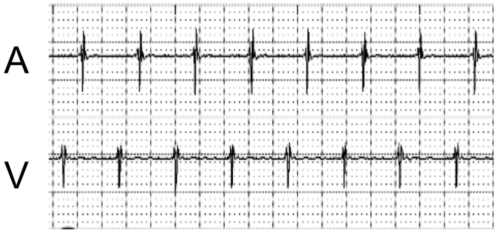
- **Ventricular Tachycardia** (**VT**): fatal; arrhythmia originates in ventricles
- **Supra-ventricular Tachycardia** (**SVT**): non-fatal; arrhythmia originates in atria



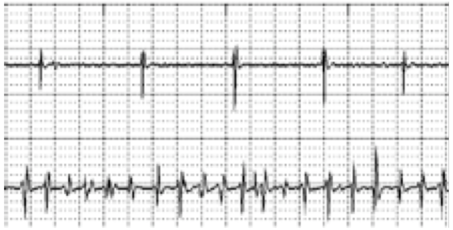
Normal sinus rhythm



Ventricular fibrillation



EGMs during SVT

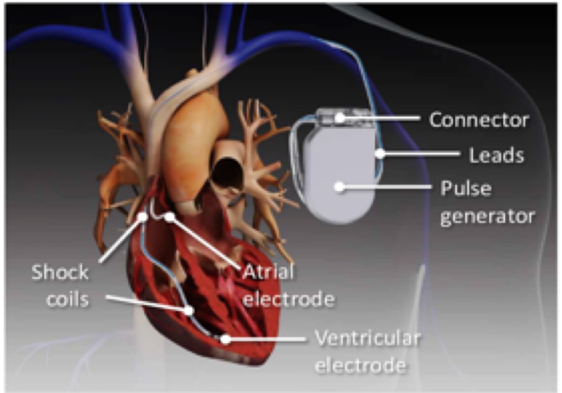


EGMs during VT

ICD communication

In-clinic settings

Patient



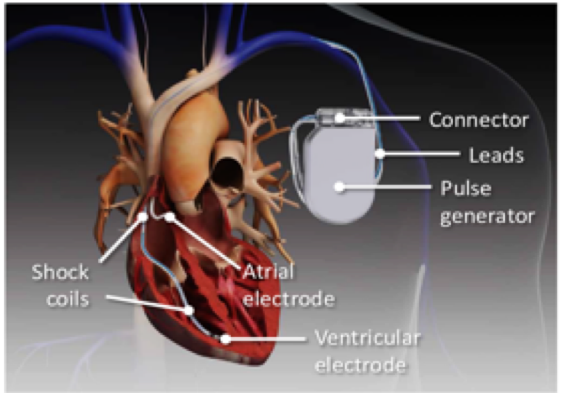
radio-frequency (RF) communication
Medical Implant Communication Service (MICS) band: 401-406 MHz

Clinician operating ICD programmer

ICD communication

In-clinic settings

Patient



Clinician operating ICD programmer

change device parameters and settings → affects discrimination algorithm and therapy

device info (model, ID), patient info, telemetry data

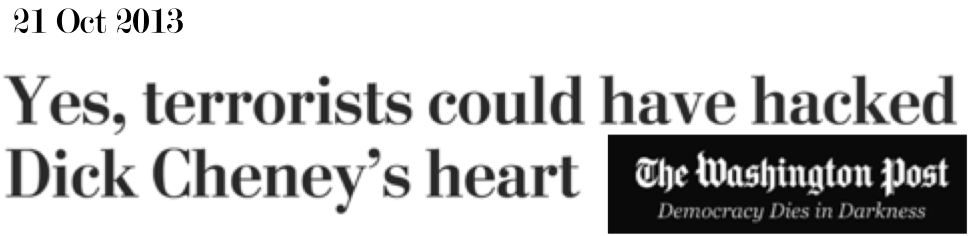
ICD communication

Remote patient monitoring – examples

Medtronic MyCareLink™ Patient monitor
Receives ICD data remotely via reader or automatically at distance (< 2m)

Medtronic MyCareLink Smart™
The reader (left) interrogates the ICD and sends medical data to smartphone app via Bluetooth

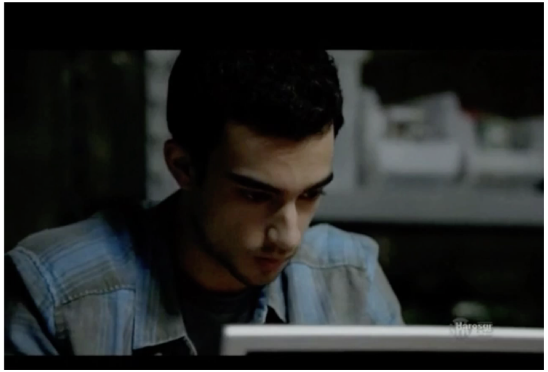
Security Concerns



21 Oct 2013

Yes, terrorists could have hacked Dick Cheney's heart

The Washington Post
Democracy Dies in Darkness



Homeland, "Broken Hearts" S2E10

Security Concerns

- ICD reprogramming attacks via software radio [Halperin et al., IEEE S&P 2008]
    - Reverse engineered devices communication protocol
    - Eavesdropping and replay (reprogramming) attacks
- ICD signal injection attacks via electromagnetic interference (EMI) [Foo Kune et al., IEEE S&P 2013]
    - EMI manipulates sensor readings by device, interrupting therapy or causing shocks
- [Aug 2017] FDA recall (firmware update) of 465,000 St Jude Medical devices to add clinician authentication

Security Concerns

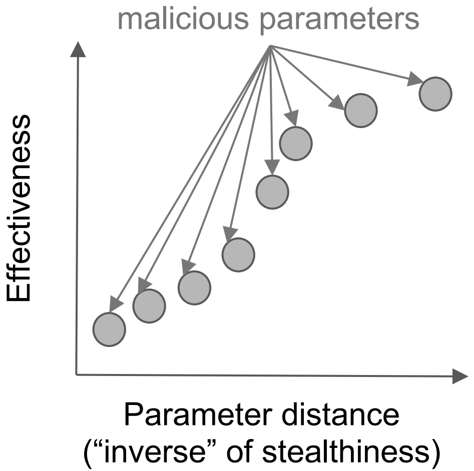
- ICD reprogramming attacks via software radio [Halperin et al., IEEE S&P 2008]
- ICD signal injection attacks via electromagnetic interference (EMI) [Foo Kune et al., IEEE S&P 2013]
- [Aug 2017] FDA recall (firmware update) of 465,000 St Jude Medical devices to add clinician authentication
- [2018-2019] Attacks on Medtronic Carelink remote monitoring system (used also for insulin pumps), exploiting absence of encryption and authentication
  - Eavesdropping, reprogramming, and also **injection of malicious programmer firmware**
  - Demonstrated by Rios and Butts at Black Hat 2018, and by researchers at Clever Security
  - US DHS issued two advisories, **with severity at 9.3/10 points** (low skill level to exploit)

Aim of this study

- ICD vulnerabilities exist, unauthorized access is possible
- **Can one reprogram an ICD to affect therapy without being detected?**
- We present a systematic method to do so

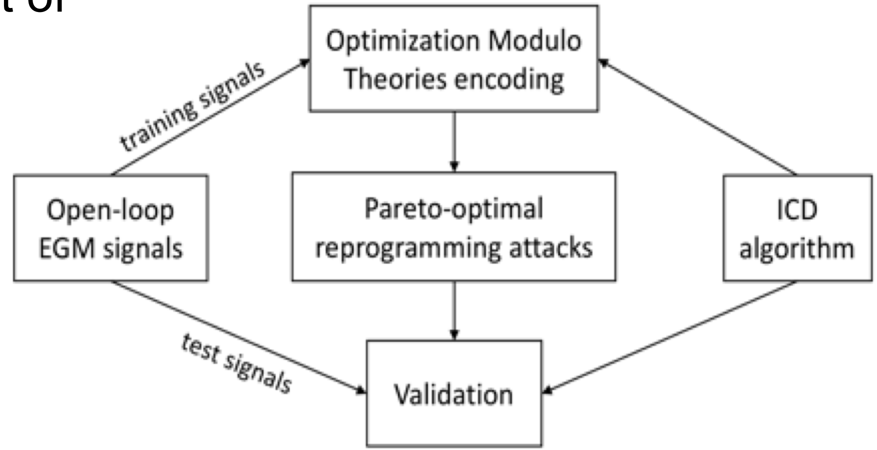
Synthesizing Stealthy Attacks on ICDs

- Reprogramming attack (manipulates ICD parameters)
- Two criteria - attack **effectiveness** and **stealthiness**
- Effectiveness:
  - Prevent necessary shocks (*fatal*)
  - Induce unnecessary shocks (*pain, tissue damage*)
- Stealthiness:
  - Attack parameters close to the nominal parameters
  - Attack should go undetected in clinical visits → small changes mistaken by clinician's error



Methodology Overview

- Synthesis as multi-objective optimization (stealthiness and effectiveness are contrasting)
  - Based on Optimization Modulo Theories (OMT) → true optima
- Model-based approach (uses a model of ICD discrimination algorithm)
- Attack effectiveness evaluated w.r.t. a set of EGM signals
- Model-based synthetic EGM signals
  - Poor availability of real patient signals
  - **Tailor attack to victim's conditions**
- Validation with unseen signals (mimics unknown victim's EGM)

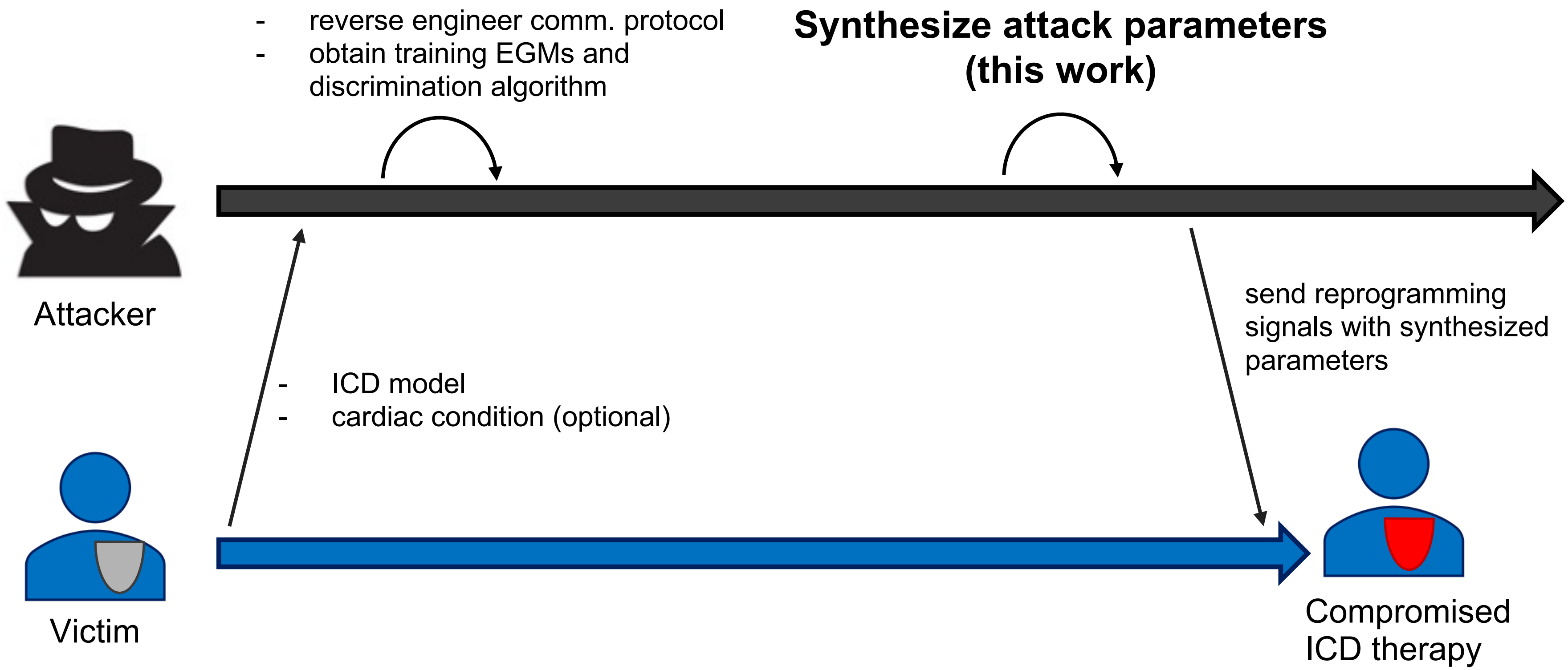

Optimization Modulo Theories encoding
training signals
Open-loop EGM signals
Pareto-optimal reprogramming attacks
ICD algorithm
test signals
Validation

Attack model

- Reprogramming: attack on **patient safety**
- Adversarial model:
  - **Active** (injects data – reprogramming commands)
  - **Unsophisticated**: must know ICD model (via discovery signals or patient records), discrimination algorithm (literature), ICD communication protocol (reverse engineering). No need for specialized equipment.
- Threat: attacker exploits **unsecure wireless interface**
- Detection mechanism: **clinician** (victim can’t monitor ICD parameters, and typically sees a doctor if the ICD doesn’t work properly)

(see [Rushanan et al, IEEE S&P 2014] for medical device security definitions)

Attack model - Timeframe



- reverse engineer comm. protocol
- obtain training EGMs and discrimination algorithm

Synthesize attack parameters (this work)

Attacker

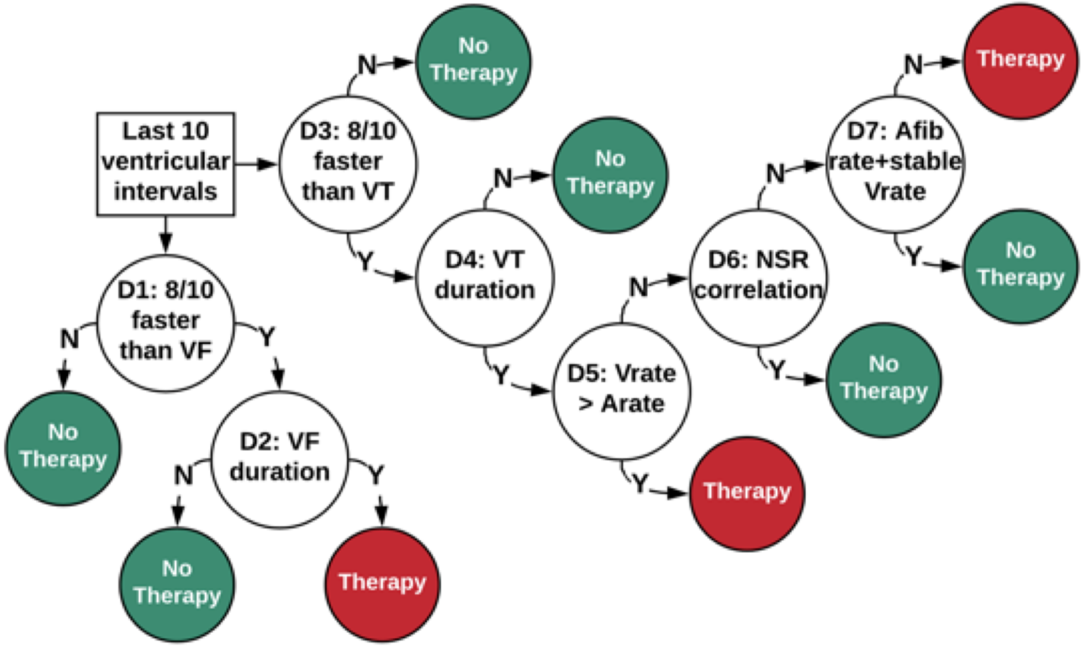
- ICD model
- cardiac condition (optional)

send reprogramming signals with synthesized parameters

Victim

Compromised ICD therapy

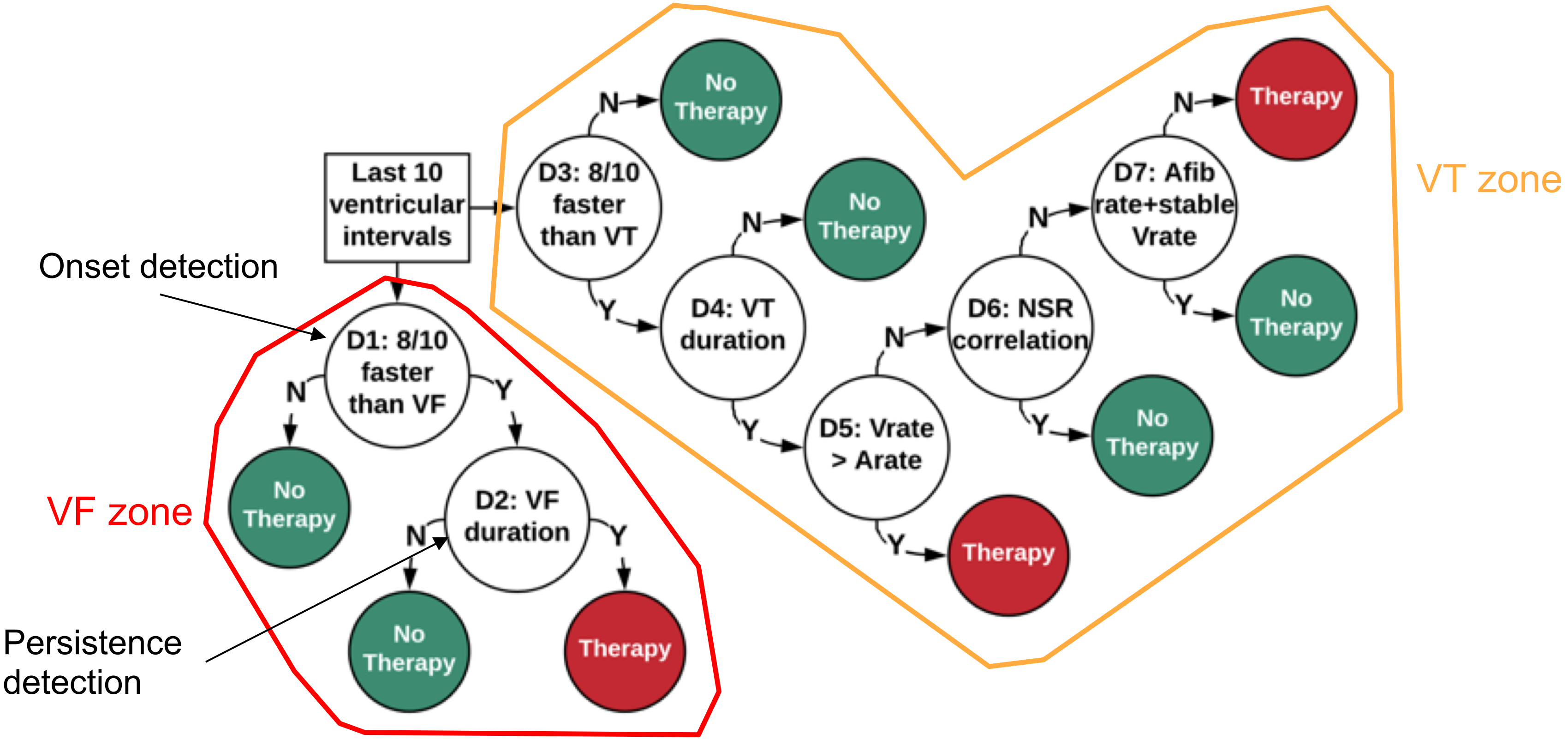
Boston Scientific ICD



BSc Rhythm ID discrimination algorithm

- Compiled from ICD manuals and medical literature by [Jiang et al, EMBC 2016]
- Conformance checked with real device in previous work

Boston Scientific ICD



Boston Scientific ICD

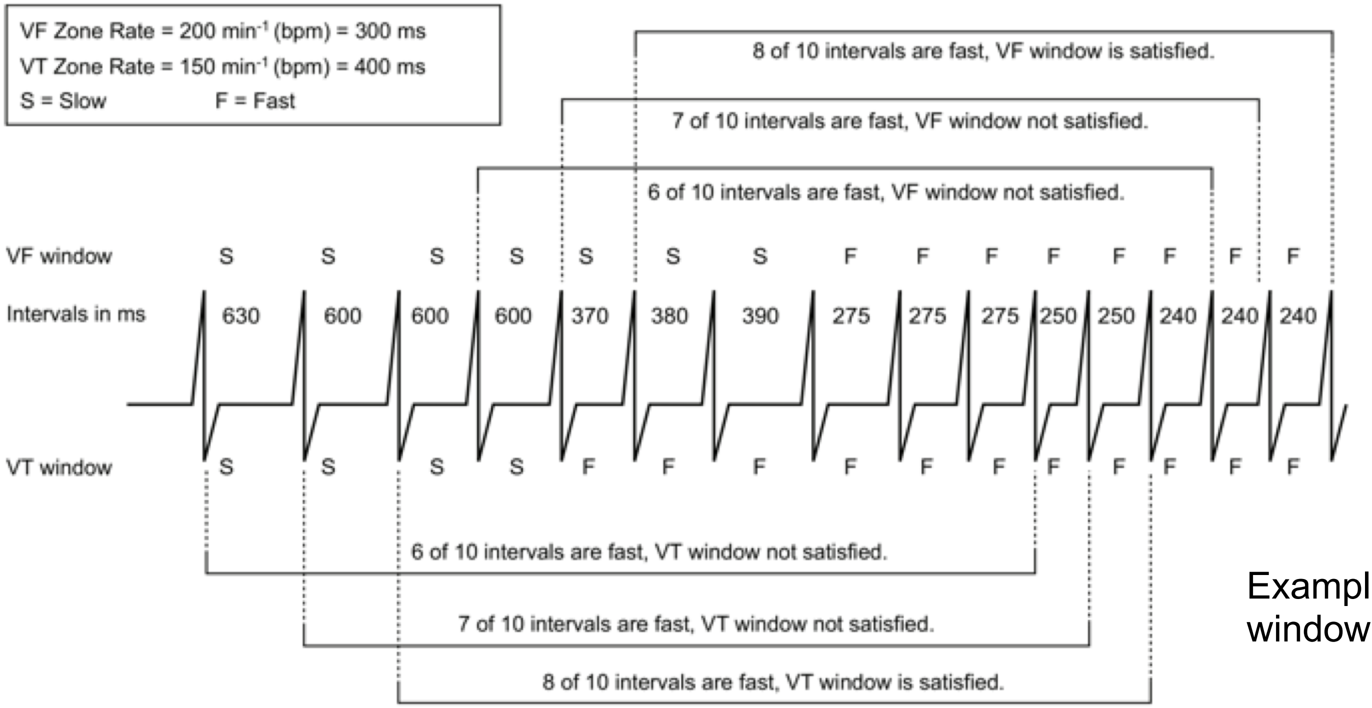

VF Zone Rate = 200 min^{-1} (bpm) = 300 ms
VT Zone Rate = 150 min^{-1} (bpm) = 400 ms
S = Slow F = Fast
8 of 10 intervals are fast, VF window is satisfied.
7 of 10 intervals are fast, VF window not satisfied.
6 of 10 intervals are fast, VF window not satisfied.
VF window
S S S S S S S F F F F F F F F
Intervals in ms
630 600 600 600 370 380 390 275 275 275 250 250 240 240 240
VT window
S S S S F F F F F F F F F F F
6 of 10 intervals are fast, VT window not satisfied.
7 of 10 intervals are fast, VT window not satisfied.
8 of 10 intervals are fast, VT window is satisfied.

Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

Example of detection windows (BSc ICD manual)

Boston Scientific ICD

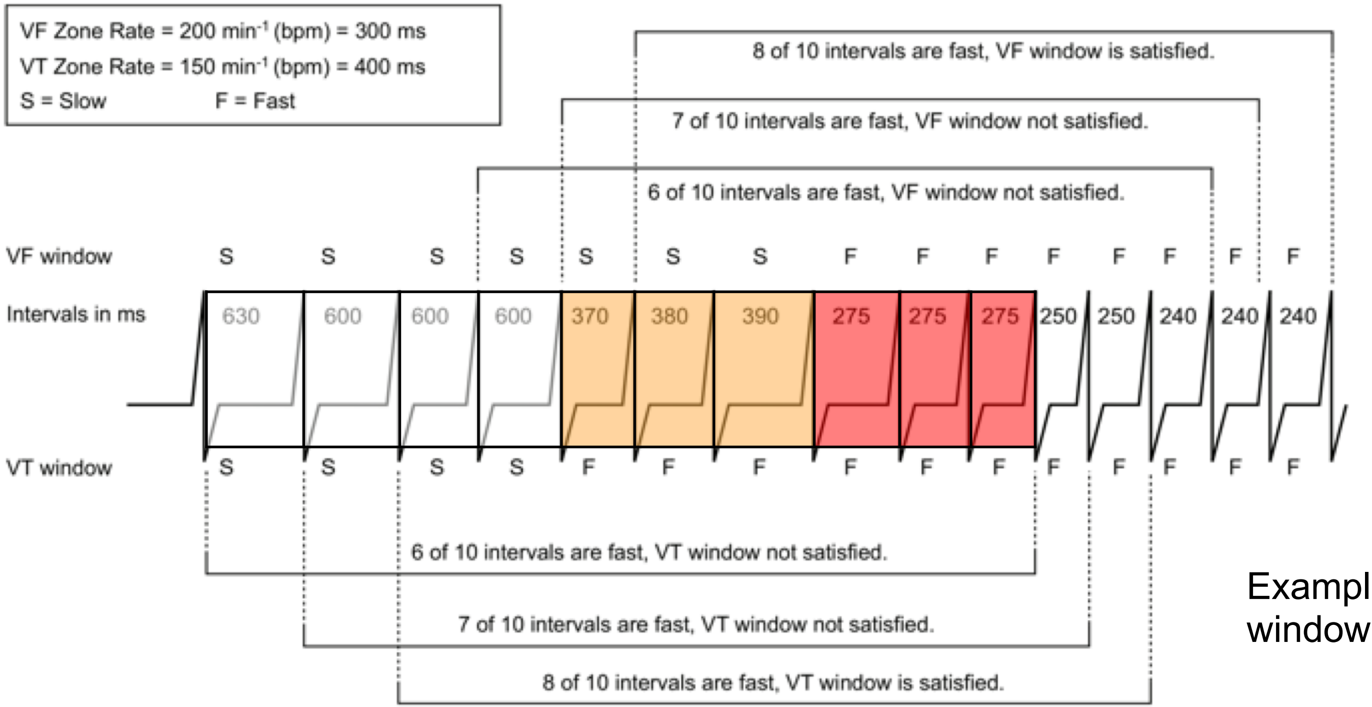

VF Zone Rate = 200 min-1 (bpm) = 300 ms
VT Zone Rate = 150 min-1 (bpm) = 400 ms
S = Slow F = Fast
8 of 10 intervals are fast, VF window is satisfied.
7 of 10 intervals are fast, VF window not satisfied.
6 of 10 intervals are fast, VF window not satisfied.
VF window
S S S S S S S F F F F F F F F
Intervals in ms
630 600 600 600 370 380 390 275 275 275 250 250 240 240 240
VT window
S S S S F F F F F F F F F F F
6 of 10 intervals are fast, VT window not satisfied.
7 of 10 intervals are fast, VT window not satisfied.
8 of 10 intervals are fast, VT window is satisfied.

Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

Faster than VT

Faster than VF

Example of detection windows (BSc ICD manual)

Boston Scientific ICD

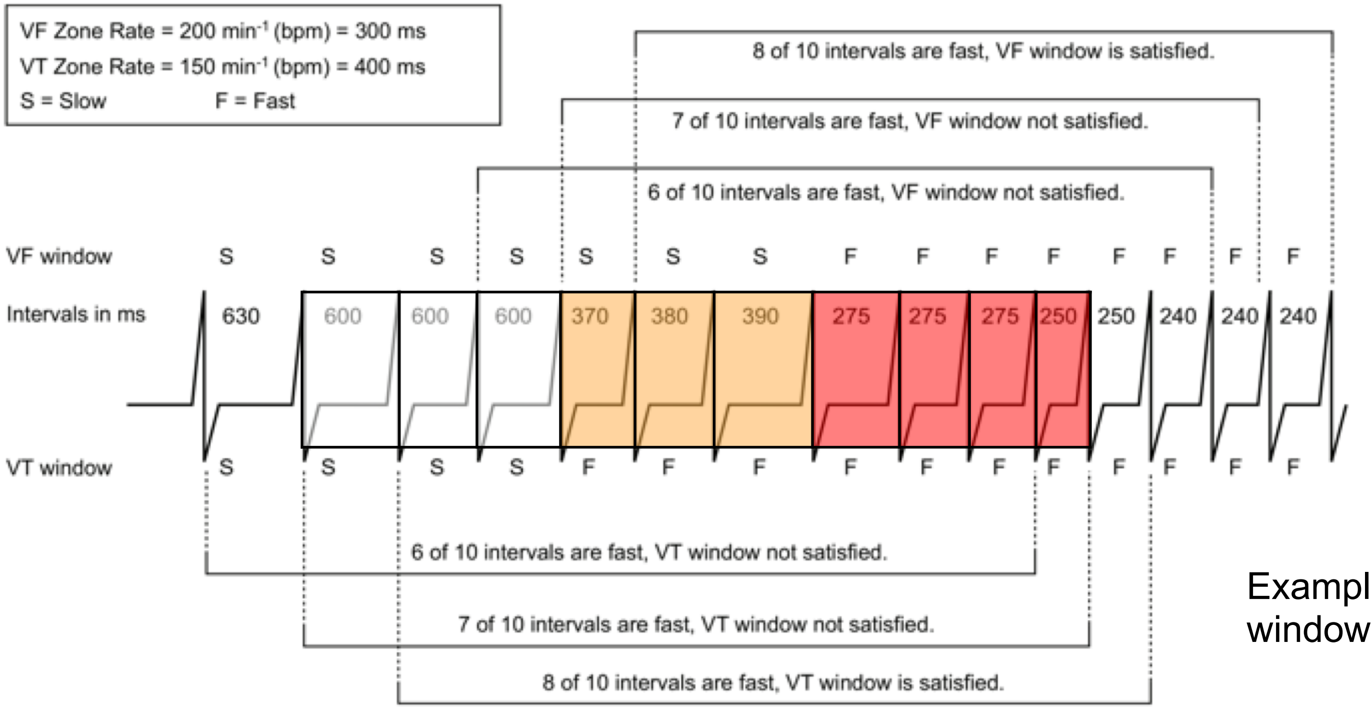
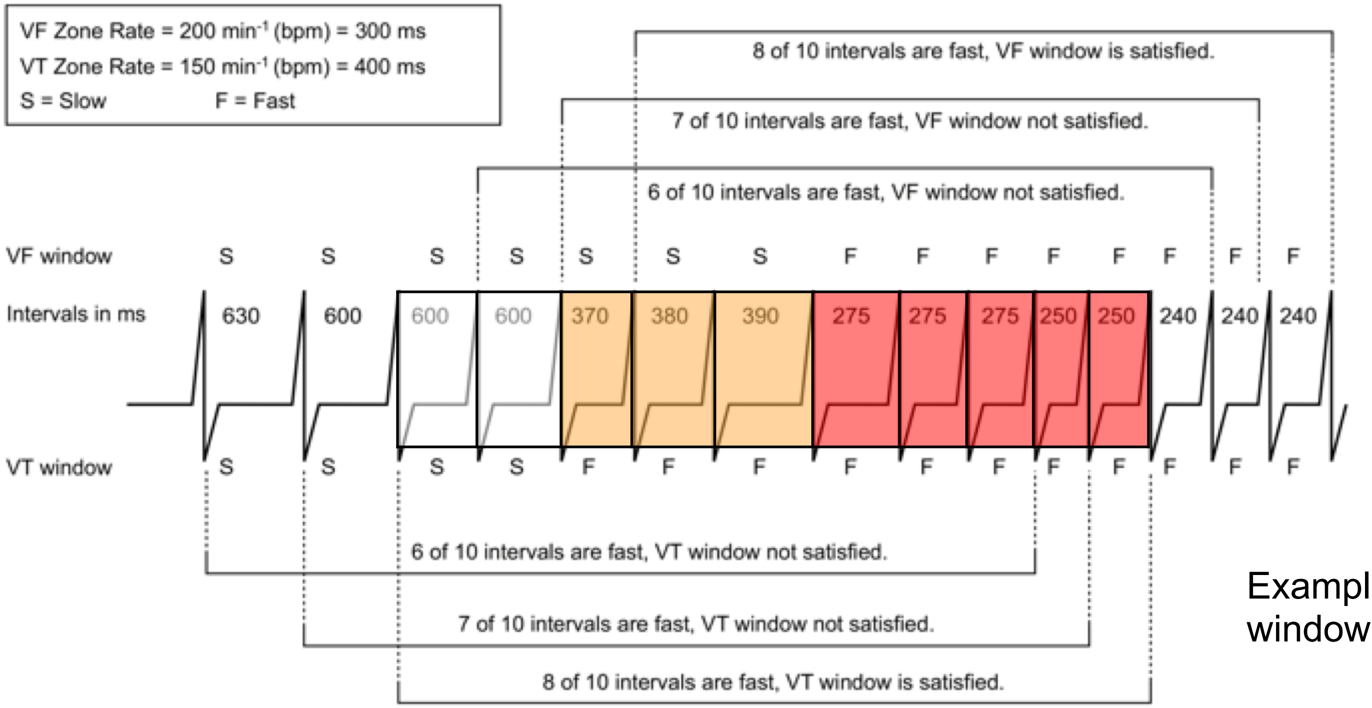

VF Zone Rate = 200 min^{-1} (bpm) = 300 ms
VT Zone Rate = 150 min^{-1} (bpm) = 400 ms
S = Slow
F = Fast
8 of 10 intervals are fast, VF window is satisfied.
7 of 10 intervals are fast, VF window not satisfied.
6 of 10 intervals are fast, VF window not satisfied.
VF window
S S S S S S S F F F F F F F F
Intervals in ms
630 600 600 600 370 380 390 275 275 275 250 250 240 240 240
VT window
S S S S F F F F F F F F F F F
6 of 10 intervals are fast, VT window not satisfied.
7 of 10 intervals are fast, VT window not satisfied.
8 of 10 intervals are fast, VT window is satisfied.

Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

🟧 Faster than VT
🟥 Faster than VF

Example of detection windows (BSc ICD manual)

Boston Scientific ICD



VF Zone Rate = 200 min^{-1} (bpm) = 300 ms
VT Zone Rate = 150 min^{-1} (bpm) = 400 ms
S = Slow F = Fast

8 of 10 intervals are fast, VF window is satisfied.

7 of 10 intervals are fast, VF window not satisfied.

6 of 10 intervals are fast, VF window not satisfied.

VF window	S	S	S	S	S	S	S	F	F	F	F	F	F	F	F
Intervals in ms	630	600	600	600	370	380	390	275	275	275	250	250	240	240	240
VT window	S	S	S	S	F	F	F	F	F	F	F	F	F	F	F

6 of 10 intervals are fast, VT window not satisfied.

7 of 10 intervals are fast, VT window not satisfied.

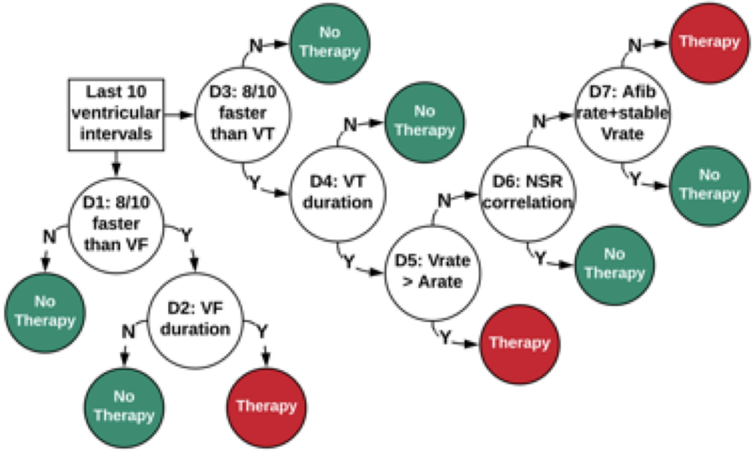
8 of 10 intervals are fast, VT window is satisfied.

Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

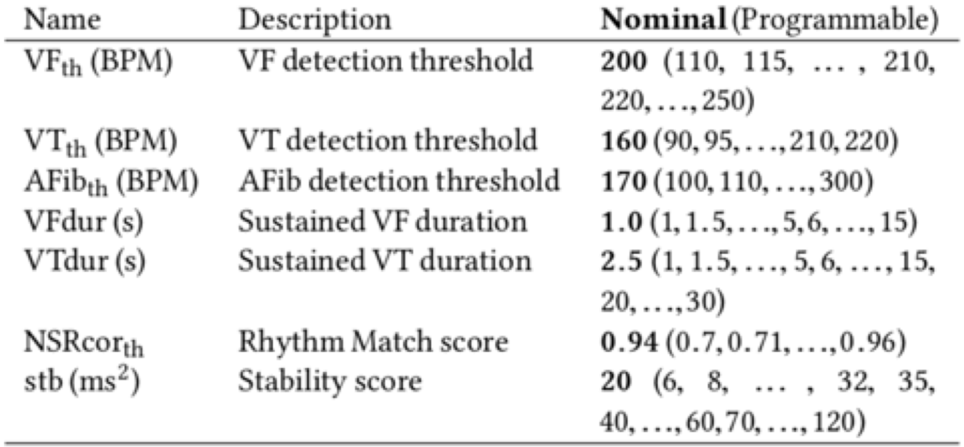
🟧	Faster than VT
🟥	Faster than VF

Example of detection windows (BSc ICD manual)

Boston Scientific ICD



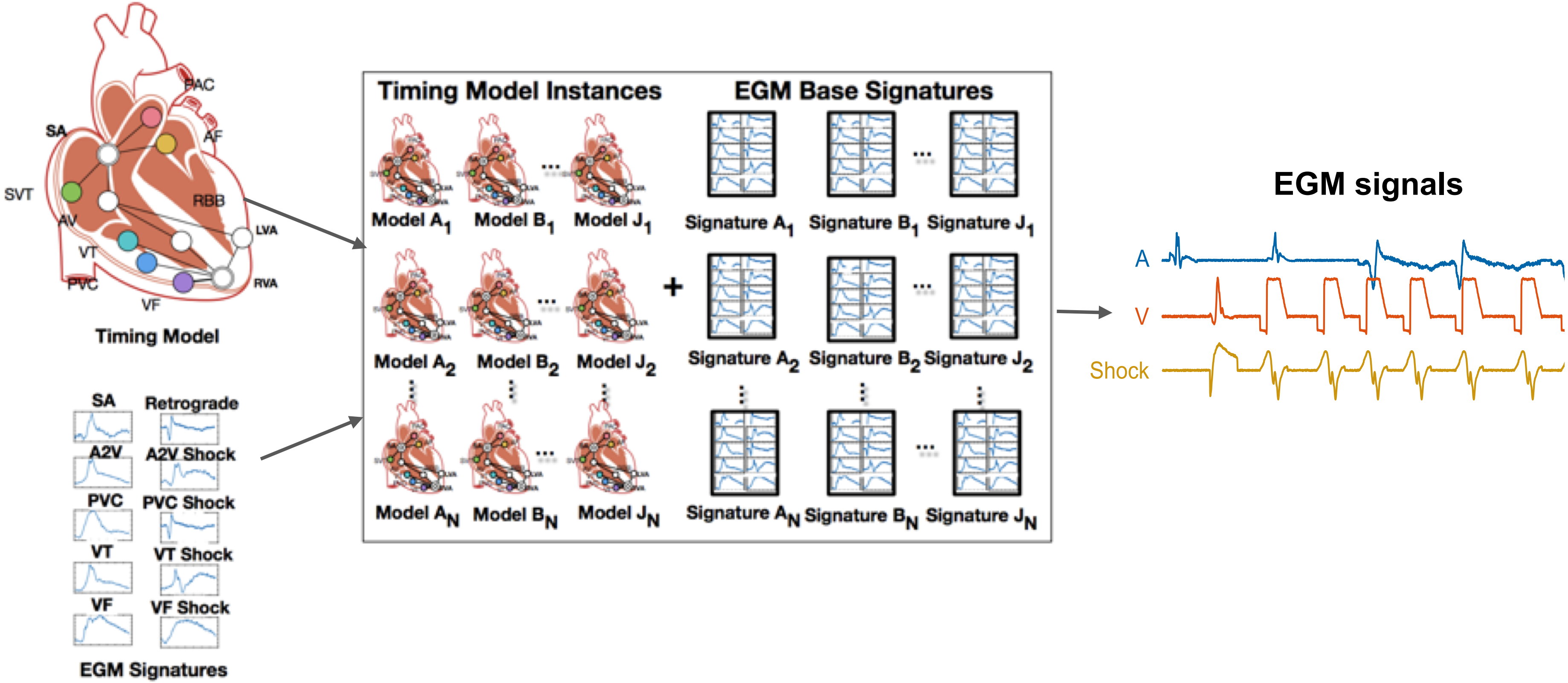
Rhythm ID discrimination algorithm



Name	Description	**Nominal** (Programmable)
VF_{th} (BPM)	VF detection threshold	**200** (110, 115, … , 210, 220, …, 250)
VT_{th} (BPM)	VT detection threshold	**160** (90, 95, …, 210, 220)
$AFib_{th}$ (BPM)	AFib detection threshold	**170** (100, 110, …, 300)
VFdur (s)	Sustained VF duration	**1.0** (1, 1.5, …, 5, 6, …, 15)
VTdur (s)	Sustained VT duration	**2.5** (1, 1.5, …, 5, 6, …, 15, 20, …, 30)
$NSRcor_{th}$	Rhythm Match score	**0.94** (0.7, 0.71, …, 0.96)
stb (ms^2)	Stability score	**20** (6, 8, … , 32, 35, 40, …, 60, 70, …, 120)

Programmable parameters

Synthetic EGM signals [Jiang et al. EMBC 2016]



Attack effectiveness

“An attack is effective on a signal if it prevents required therapy or introduces inappropriate therapy”

$$f_e(\boxed{\mathbf{p}}, \boxed{S}) = \frac{1}{|S|} \cdot \sum_{\mathbf{s} \in S} I\left(\boxed{R_{th}(d,\mathbf{p},\mathbf{s})} \neq \boxed{R_{th}(d,\mathbf{p}^*,\mathbf{s})}\right)$$

Attack parameters

Set of signals (training or test)

True iff therapy is given at any point in signal **s** under attack parameters **p**

True iff therapy is given at any point in **s** under nominal parameters **p***

Attack effectiveness (example)

	Therapy signal with nominal parameters $d(\mathbf{p}^*)(\mathbf{s})$									Therapy signal with attack parameters $d(\mathbf{p})(\mathbf{s})$								Effective?
$\mathbf{s} = \mathbf{s}_1$	☐	☐	🟥	☐	☐	🟥	☐	☐	$\longrightarrow$	☐	☐	☐	☐	☐	🟥	☐	☐	✗
$\mathbf{s} = \mathbf{s}_2$	☐	☐	☐	☐	☐	☐	☐	🟥	$\longrightarrow$	☐	☐	☐	☐	☐	☐	☐	☐	✓
$\mathbf{s} = \mathbf{s}_3$	☐	☐	☐	☐	☐	☐	☐	☐	$\longrightarrow$	☐	☐	☐	☐	🟥	☐	☐	🟥	✓
$\mathbf{s} = \mathbf{s}_4$	🟥	☐	☐	☐	☐	☐	☐	☐	$\longrightarrow$	☐	☐	☐	☐	☐	☐	☐	🟥	✗
$k =$	0	1	2	3	4	5	6	7		0	1	2	3	4	5	6	7	

Heart cycles

Heart cycles

🟥 Therapy

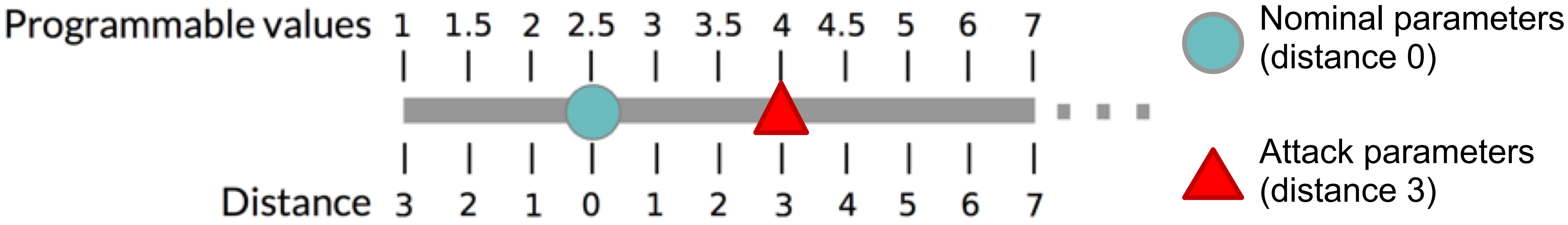
☐ No therapy

Attack stealthiness

"An attack is stealthy when the deviation from the nominal parameters is small"

We quantify stealthiness as parameter distance (number of programmable values separating nominal and attack parameters – max separation over all parameters)

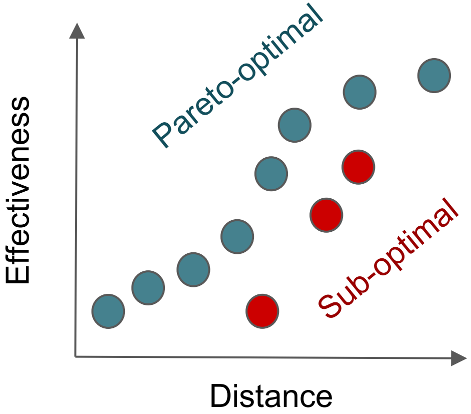
Example: parameter VT duration (s)


Programmable values
1
1.5
2
2.5
3
3.5
4
4.5
5
6
7
Distance
3
2
1
0
1
2
3
4
5
6
7
Nominal parameters (distance 0)
Attack parameters (distance 3)

Synthesis of optimal stealthy attacks

Derive the set **P** of Pareto-optimal ICD parameters wrt effectiveness f_e and distance f_s objectives

$$\mathbf{P} = \{\mathbf{p} \in \mathbb{P} \mid \nexists \mathbf{p}' \in \mathbb{P}.\, (f_e(\mathbf{p}', S) > f_e(\mathbf{p}, S) \wedge f_s(\mathbf{p}') \leq f_s(\mathbf{p})) \vee (f_e(\mathbf{p}', S) \geq f_e(\mathbf{p}, S) \wedge f_s(\mathbf{p}') < f_s(\mathbf{p}))\}$$



Solution technique - optimization modulo theories (OMT)

- Our optimization problem is challenging
  - nonlinear, non-convex, combinatorial, constrained by ICD algorithm

- SMT (SAT + theories) is well-suited to solve combinatorial problems
  [De moura and Bjorner, CACM Sep 2011]

- **SMT encoding of BSc ICD algorithm:**
  - formalization as a set FOL formulas over decidable theories (SMT QF_LIRA)
  - **Efficient encoding:** signal processing (e.g. peak detection) and nonlinear operations (e.g. correlation scores) not dependent on ICD parameters are precomputed
  - Parameter synthesis = finding a model, i.e., a SAT assignment of variables

Solution technique - optimization modulo theories (OMT)

- **SMT encoding of BSc ICD algorithm:**
    - formalization as a set FOL formulas over decidable theories (SMT QF_LIRA)
    - **Efficient encoding:** signal processing (e.g. peak detection) and nonlinear operations (e.g. correlation scores) not dependent on ICD parameters are precomputed
    - Parameter synthesis = finding a model, i.e., a SAT assignment of variables

- **OMT = SMT + precise optimization**
  [Bjørner et al., TACAS 2015, Sebastiani et al., CAV 2015]
    - find the model (among all SAT assignments) that optimizes some objectives
    - Guided improvement algorithm for multi-objective optimization
      [Rayside et al, MIT-CSAIL-TR-2009-033]

SMT encoding (intuition)

BMC-like formulation:

[Biere et al, TACAS 1999]

$$\boxed{\text{paramRanges}} \wedge \bigwedge_{j=1}^{|S|} \left(\boxed{\text{Init}(s_{j,0})} \wedge \bigwedge_{k=0}^{N_j-1} \boxed{T(k, s_{j,k}, s_{j,k+1})} \right)$$

Constraints for programmable ranges

Initial state of ICD algorithm on j-th signal

Unrolling of transition relation describing evolution of the ICD state between heart cycles

ICD state for j-th signal and k-th heart cycle:

$$s_{j,k} \overset{\text{def}}{=} (\boxed{\text{VFd}_{j,k}}, \boxed{\text{VTd}_{j,k}}, \boxed{\text{tVF}_{j,k}}, \boxed{\text{tVT}_{j,k}}) \in \mathbb{B} \times \mathbb{B} \times \mathbb{Z}^{\geq} \times \mathbb{Z}^{\geq}$$

In VF duration?

In VT duration?

Time spent in VFd

Time spent in VTd

SMT encoding (intuition)

Transition function:

$$((\neg \mathrm{VFd}_{k} \wedge \neg \mathrm{VFstart}_{k}) \Rightarrow \neg \mathrm{VFd}_{k+1})$$

“If outside VF duration and no VF episodes are detected, then stay outside VF duration in the next state”

$$((\mathrm{VFstart}_{k} \wedge (\neg \mathrm{VFd}_{k} \vee \mathrm{VFend}_{k})) \Rightarrow \mathrm{VFd}_{k+1})$$

“If a VF episode is detected and we are outside VF duration or VF duration just ended, then enter VF duration in the next state”

...

Full encoding available in [Paoletti et al, arXiv:1810.03808]

SMT encoding (intuition)

$$s_{j,k} \overset{\textbf{def}}{=} (\boxed{\mathrm{VFd}_{j,k}}, \boxed{\mathrm{VTd}_{j,k}}, \boxed{\mathrm{tVF}_{j,k}}, \boxed{\mathrm{tVT}_{j,k}}) \in \mathbb{B}\times\mathbb{B}\times\mathbb{Z}^{\geq}\times\mathbb{Z}^{\geq}$$

In VF duration?

In VT duration?

Time spent in VFd

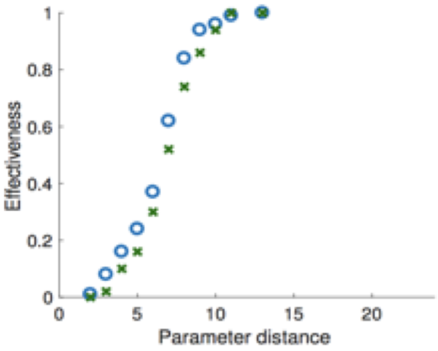
Time spent in VTd

$$\ldots(\bot,\bot,0,0) \xrightarrow{13} (\bot,\top,0,0) \xrightarrow{14} (\bot,\top,0,309)\ldots$$

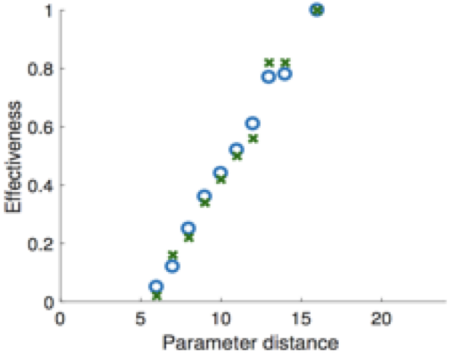
$$\xrightarrow{25} (\bot,\top,0,2317) \xrightarrow{26} (\bot,\bot,0,0)$$

Evaluation, condition-specific attacks

- Use synthetic EGMs for 19 heart conditions
  - 100 EGMs for training (synthesis), 50 EGMs for validation (per condition)



Condition 10 (VT-like)



Condition 17 (VT-like)

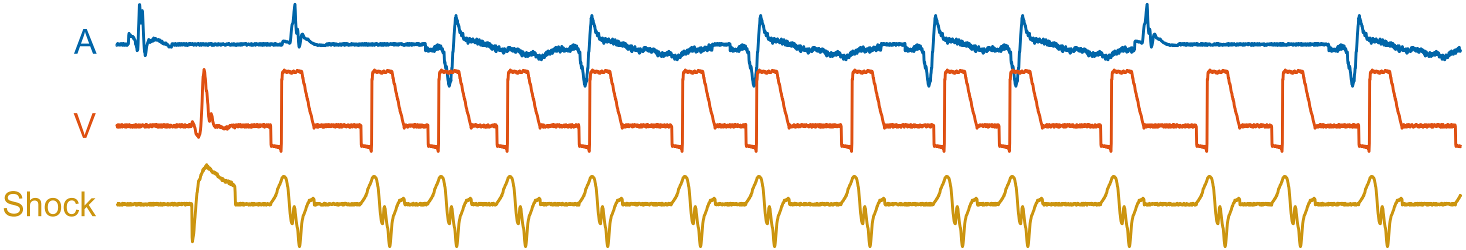
○ Training signals ✖ Validation signals

- Attacks on VT-like conditions are all very effective
- But not all equally stealthy (see left)

Common attack strategy:

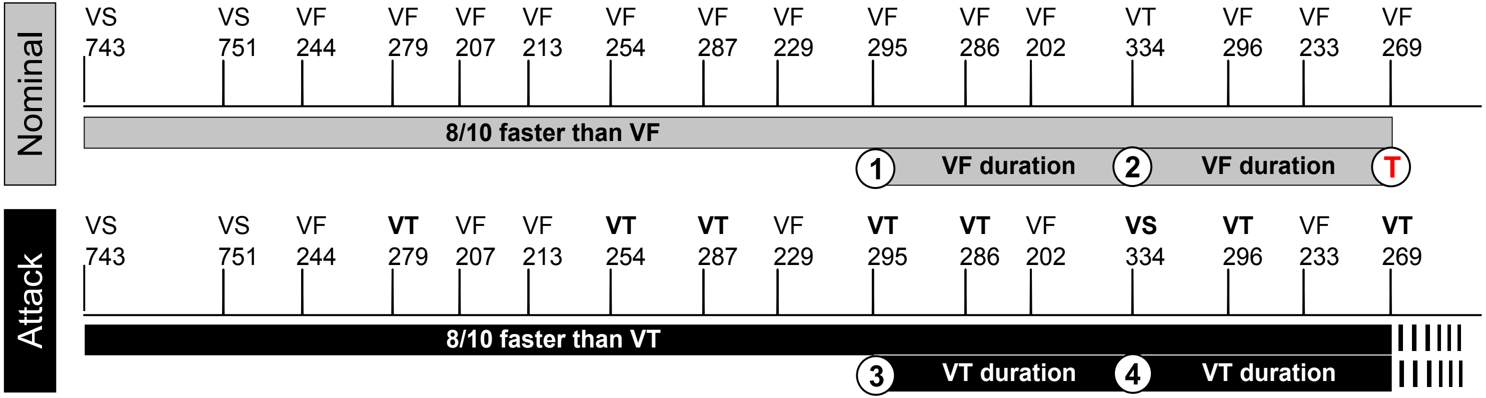
- Increase VT and VF detection thresholds in order to miss episodes
- Increase VF and VT durations to reduce probability that episode is marked sustained

Evaluation, condition-specific attacks



VF_th = 200 BPM
VT_th = 160 BPM
VFdur = 1 s
VTdur = 2.5 s

VF_th = 240 BPM
VT_th = 185 BPM
VFdur = 4 s
VTdur = 7 s



Nominal	VS 743	VS 751	VF 244	VF 279	VF 207	VF 213	VF 254	VF 287	VF 229	VF 295	VF 286	VF 202	VT 334	VF 296	VF 233	VF 269
Attack	VS 743	VS 751	VF 244	**VT** 279	VF 207	VF 213	**VT** 254	**VT** 287	VF 229	**VT** 295	**VT** 286	VF 202	**VS** 334	**VT** 296	VF 233	**VT** 269

8/10 faster than VF

1 **VF duration** **2** **VF duration** **T**

8/10 faster than VT

3 **VT duration** **4** **VT duration**

EGM extract from condition 10 signals

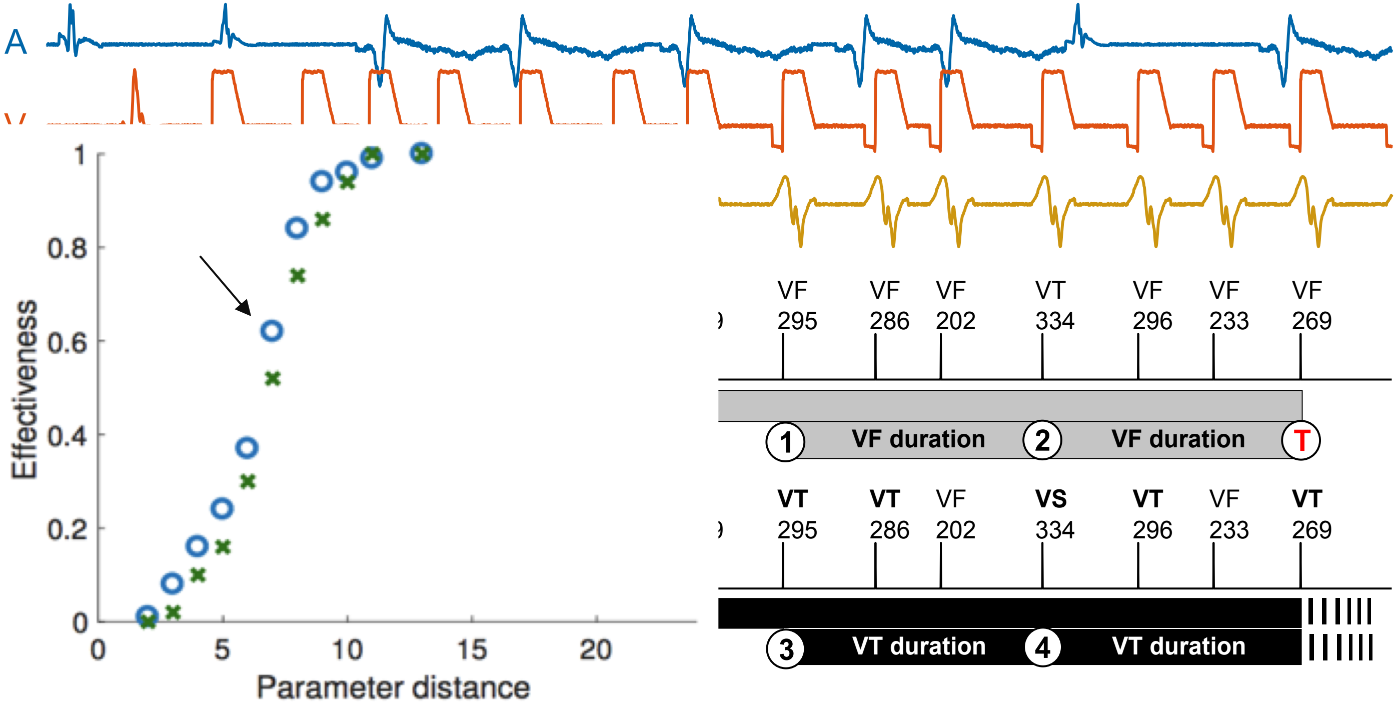
Evaluation, condition-specific attacks

VF_th = 200 BPM
VT_th = 160 BPM
VFdur = 1 s
VTdur = 2.5 s

Nominal

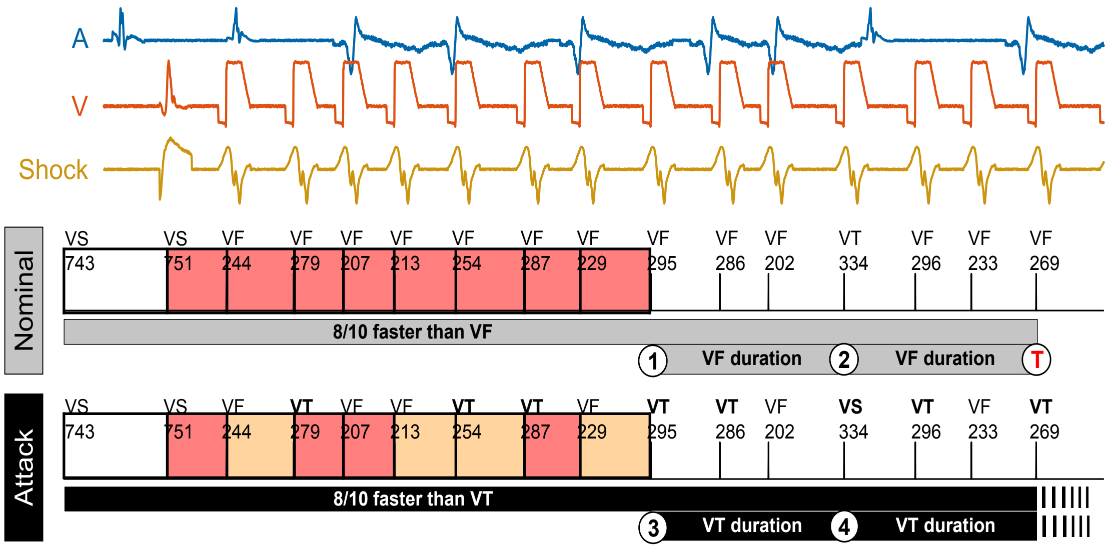
VF_th = 200 BPM
VT_th = 160 BPM
VFdur = 4 s
VTdur = 7 s

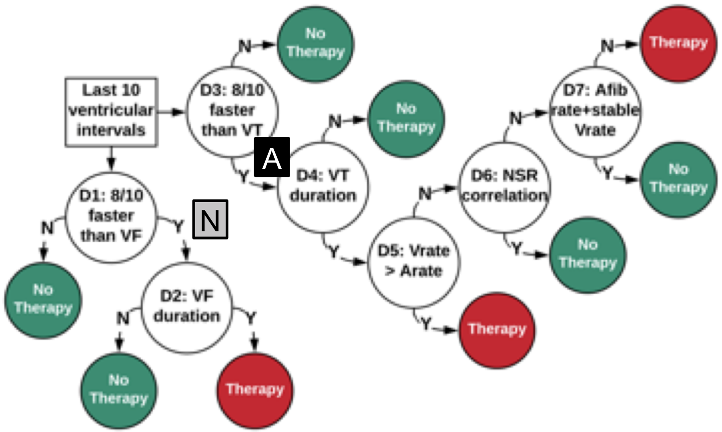
Attack



EGM extract from condition 10 signals

Evaluation, condition-specific attacks

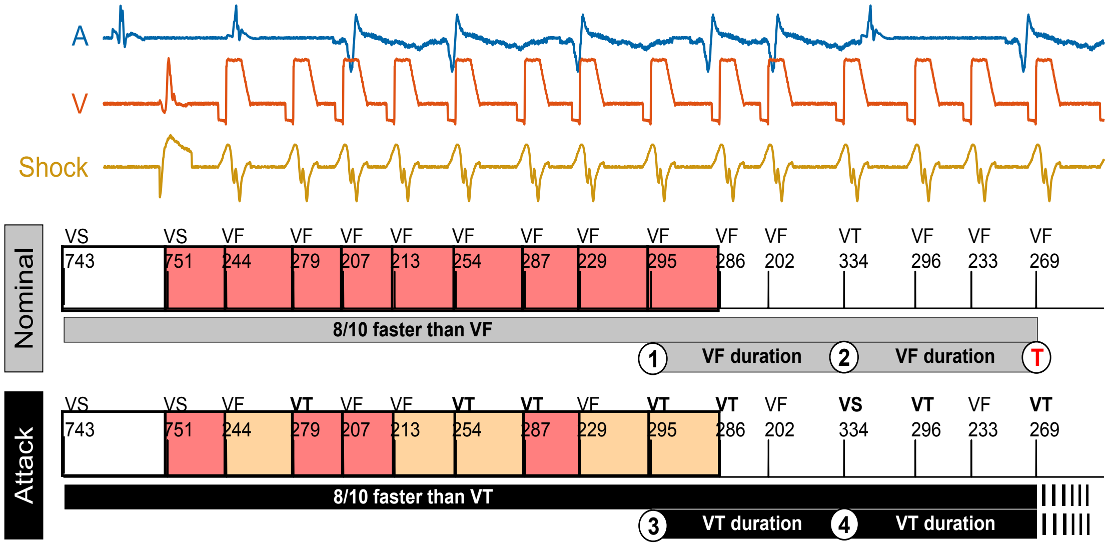


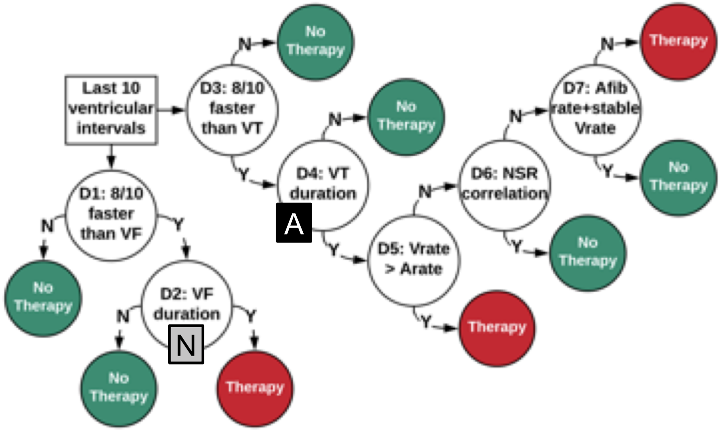


Faster than VT

Faster than VF

Evaluation, condition-specific attacks

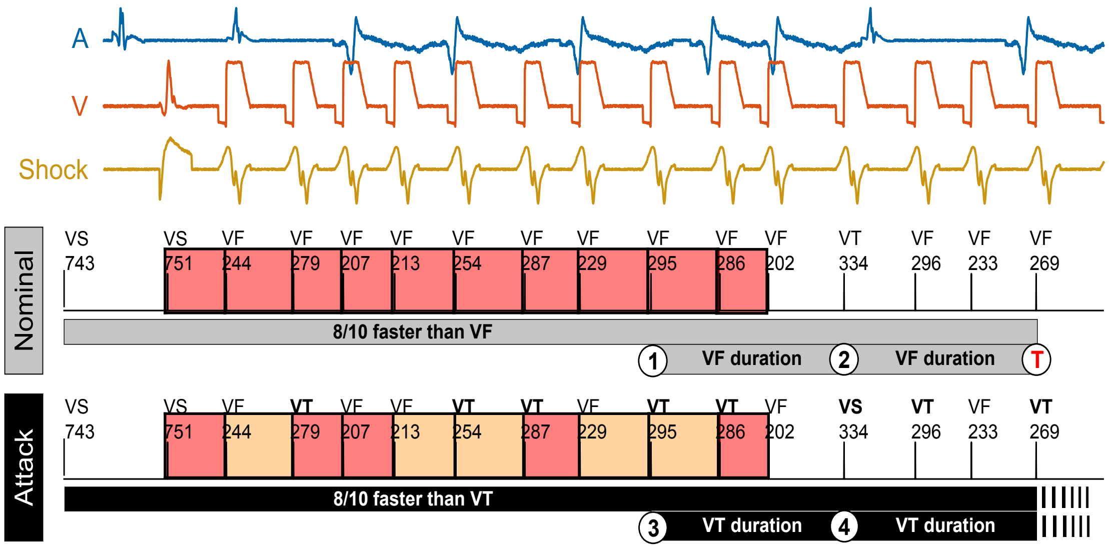


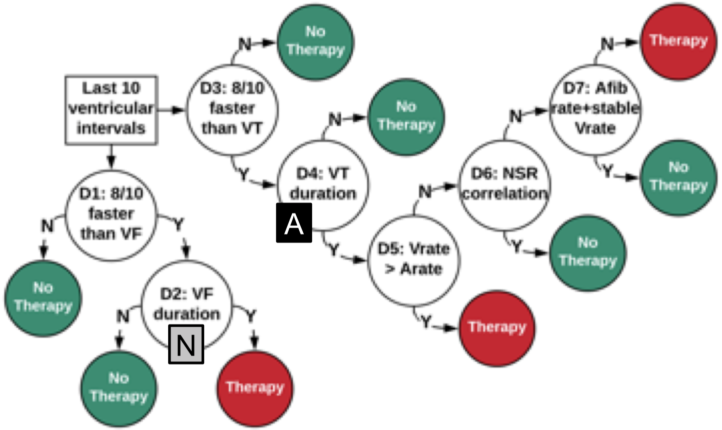


🟧 Faster than VT

🟥 Faster than VF

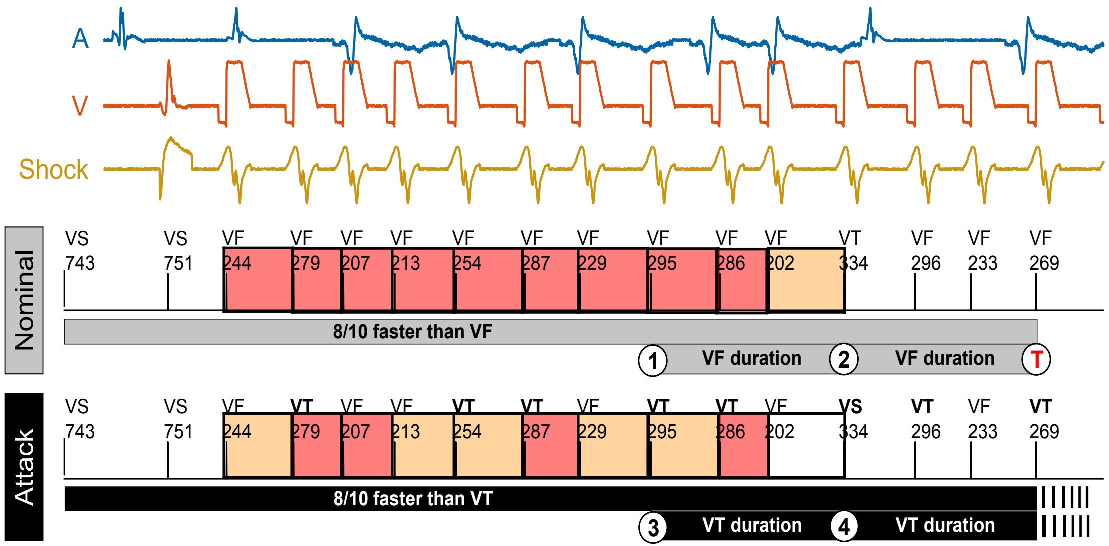
Evaluation, condition-specific attacks

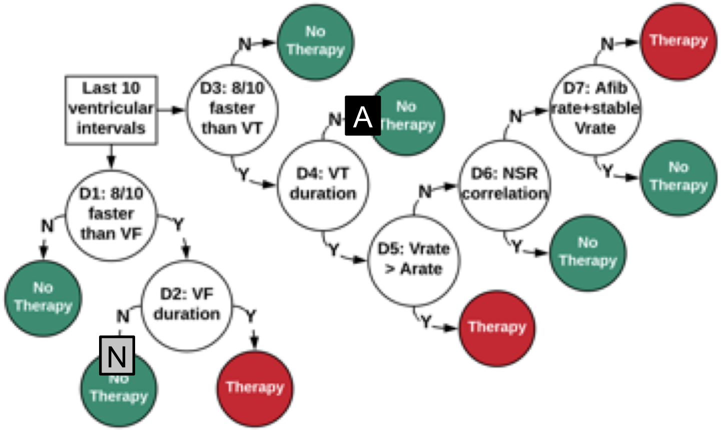




Faster than VT
Faster than VF

Evaluation, condition-specific attacks

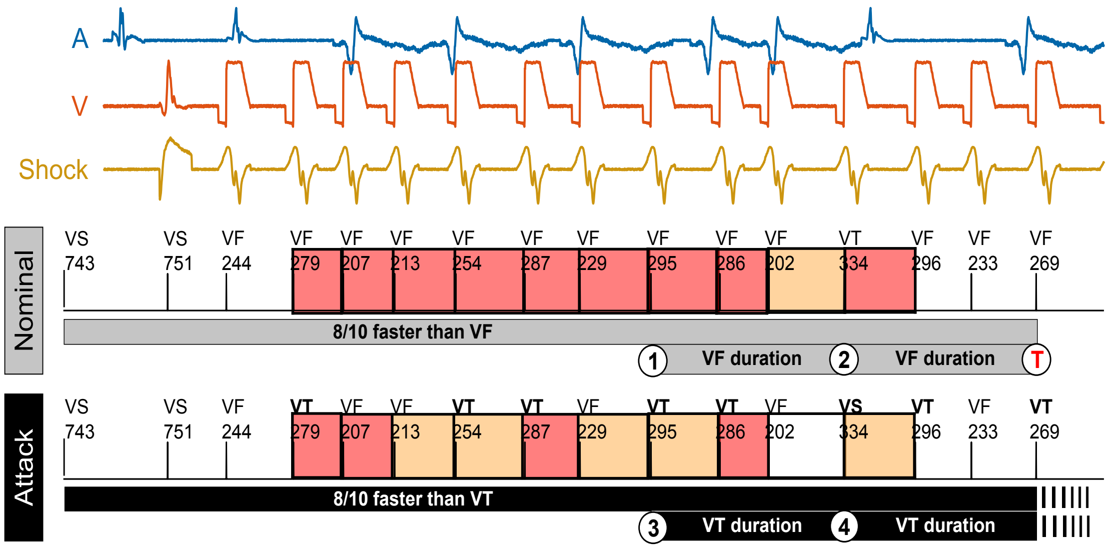


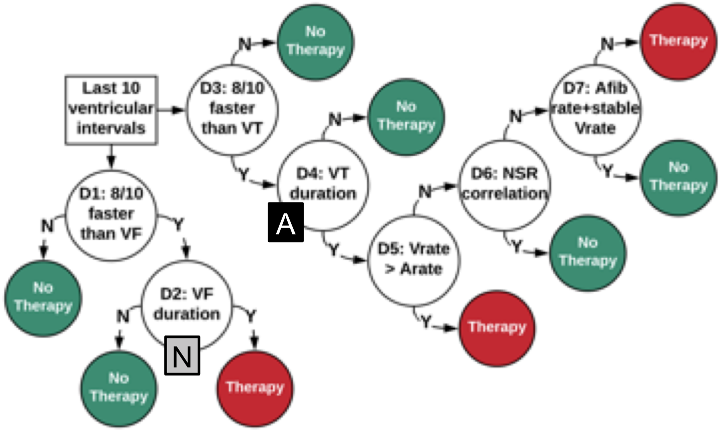


Faster than VT

Faster than VF

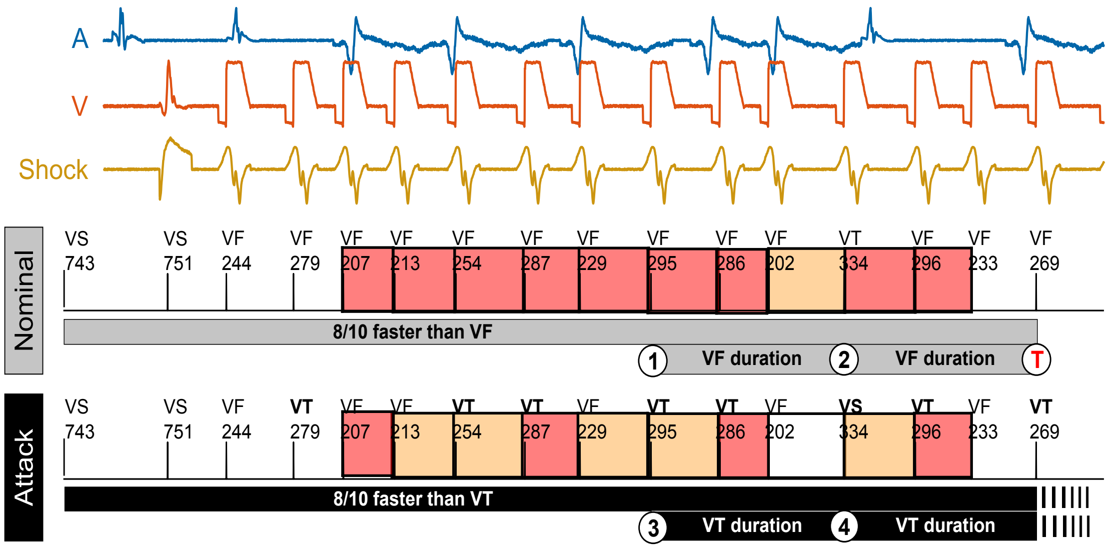
Evaluation, condition-specific attacks

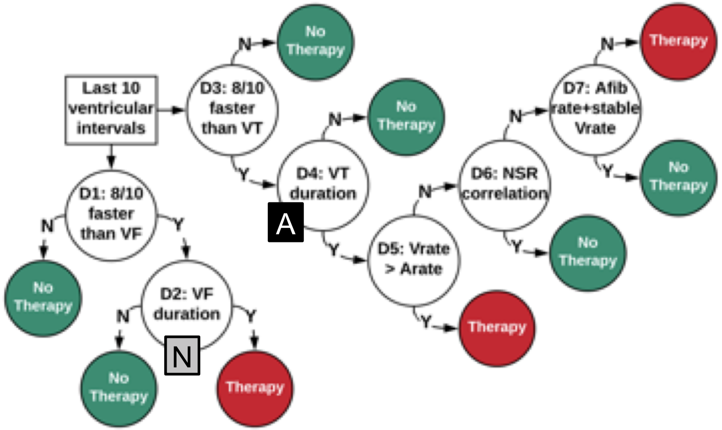




Faster than VT
Faster than VF

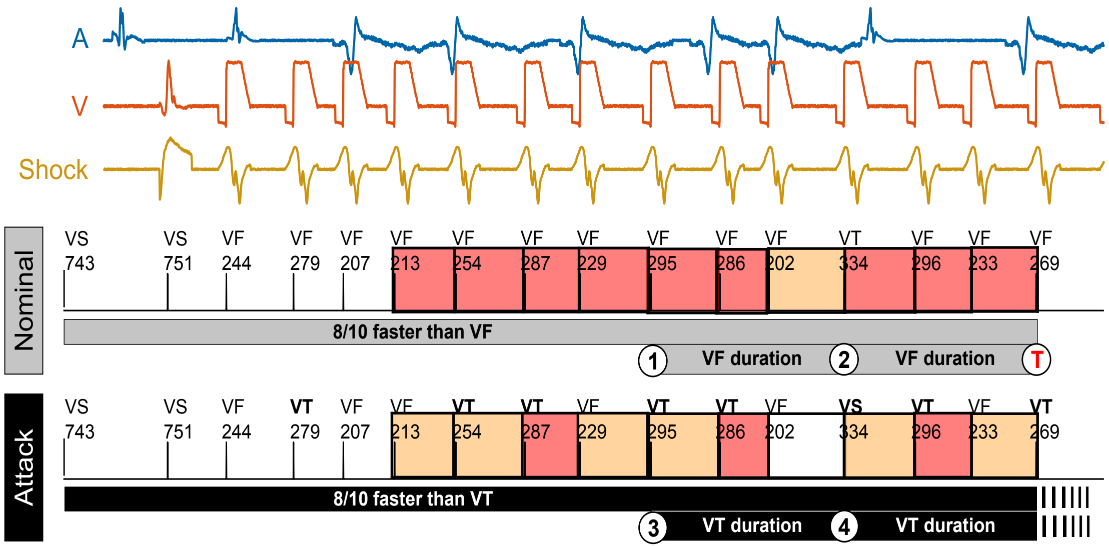
Evaluation, condition-specific attacks

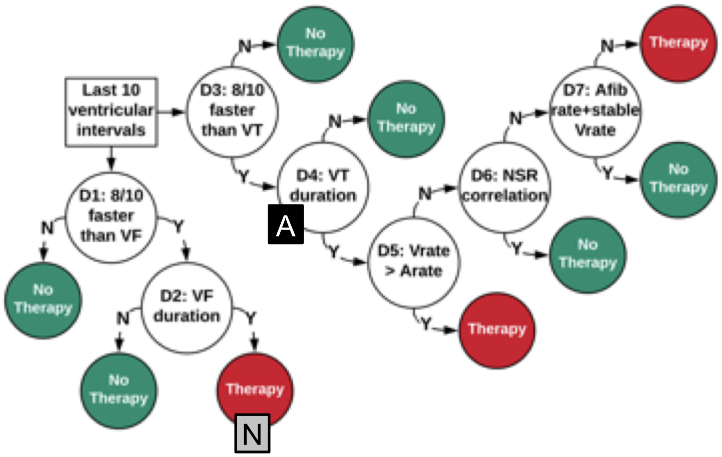




🟧	Faster than VT
🟥	Faster than VF

Evaluation, condition-specific attacks



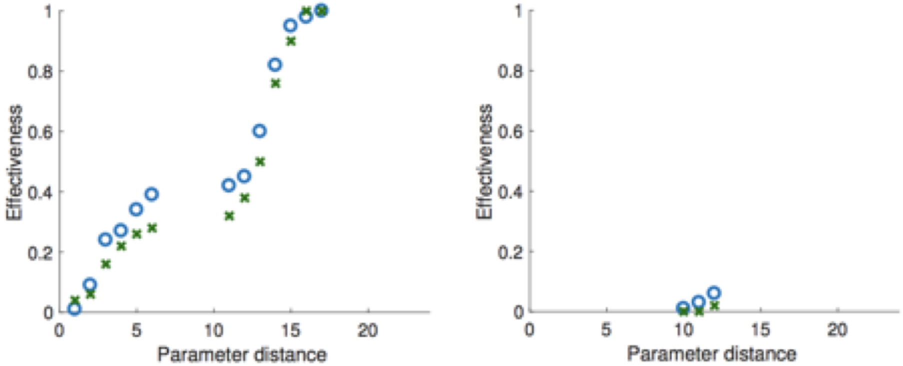


Therapy prevented by attack

Faster than VT

Faster than VF

Evaluation, condition-specific attacks



Condition 5 (SVT-like)

Condition 11 (SVT-like)

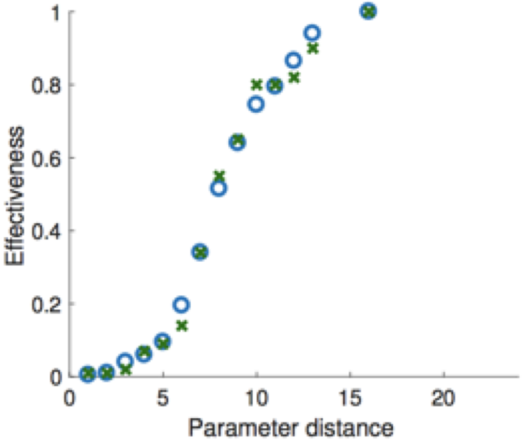
○ Training signals

✖ Validation signals

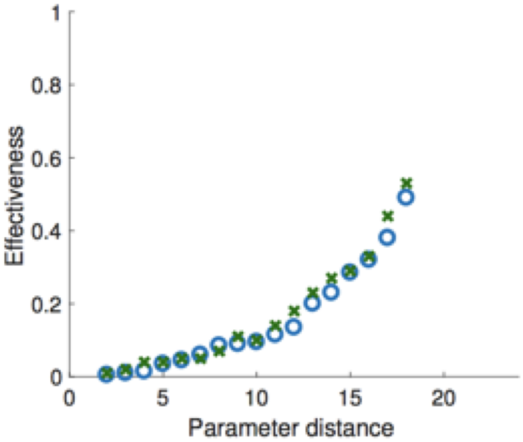
- Attacks on SVT-like conditions are not all equally effective
- Because, under normal HR, VT and VF must be reprogrammed to very low values to classify it as fast HR
- Common attack strategy: keep VF/VT thresholds and duration to a minimum

Evaluation, condition-agnostic attacks

- Two groups of signals obtained by merging VT-like and SVT-like EGMs
  - Useful when the attacker has little knowledge of the victim
  - 200 EGMs for training, 100 EGMs for validation



VT-like conditions



SVT-like conditions

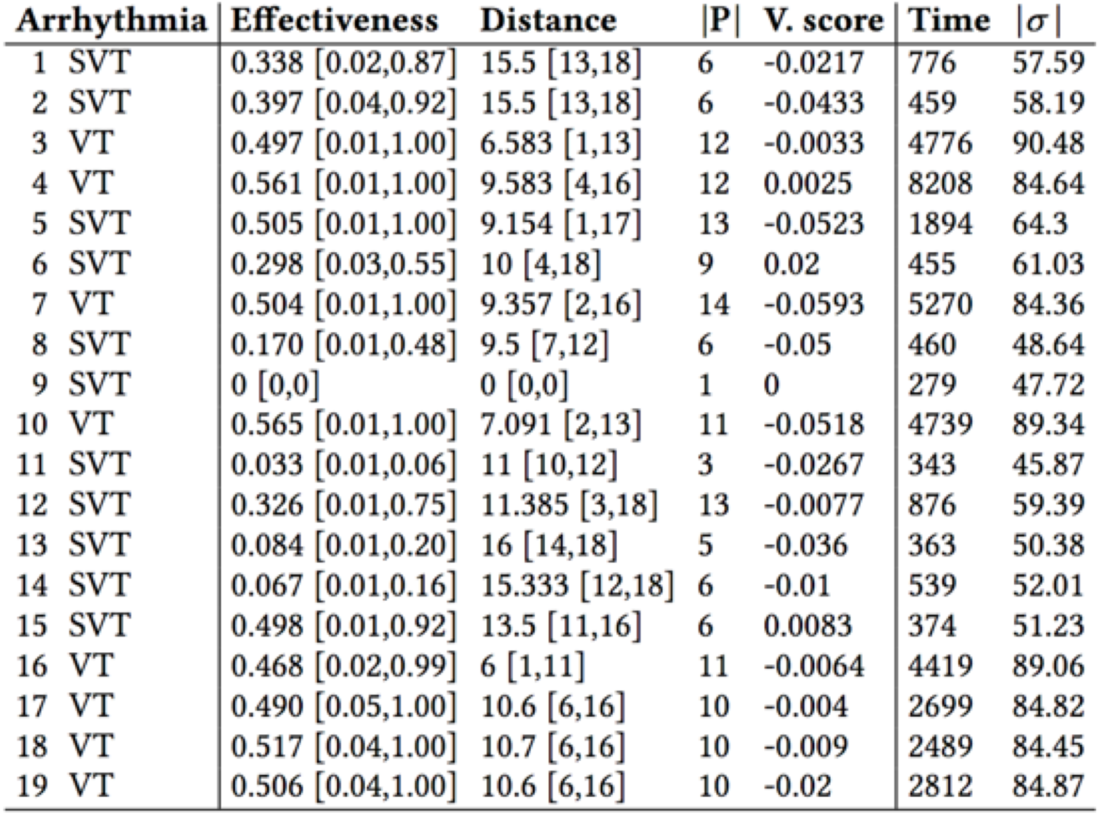
Countermeasures

- Secure authentication with key generated from patient biometrics (ECG)
  [Xu et al, IEEE InfoCom 2011, …]
- Distance-bounding protocols, to allow communication only at short distances
  [Rasmussen et al, CCS 2009,…]
- External “mediator” device: authenticates with both device and programmer, thus protecting against unauthorized communication
  [Denning et al, HotSec’08,…]
- Attack detection via ICD beeping on communication
  [Halperin et al, IEEE S&P 2008]
- Store copy of “true” parameters in both hospital DB and ICD, and regularly check for consistence

Conclusion

- Attacks on cardiac devices are a serious threat, exploiting unsecure wireless communication
- We presented the first method to synthesize stealthy reprogramming attacks tailored to the victim’s conditions
- Employs synthetic EGMs and automated reasoning (OMT) to find malicious parameters with optimal effectiveness-stealthiness trade-offs
- Well generalizes to unseen data (mimicking unknown victim EGM)
- **Future work:** evaluation on real ICD, other ICD models, real patient EGMs, closed-loop interaction, synthesis of robust discrimination algorithms

Statistics of condition-specific attacks



Arrhythmia	Effectiveness	Distance	$\|P\|$	V. score	Time	$\|\sigma\|$
1 SVT	0.338 [0.02,0.87]	15.5 [13,18]	6	-0.0217	776	57.59
2 SVT	0.397 [0.04,0.92]	15.5 [13,18]	6	-0.0433	459	58.19
3 VT	0.497 [0.01,1.00]	6.583 [1,13]	12	-0.0033	4776	90.48
4 VT	0.561 [0.01,1.00]	9.583 [4,16]	12	0.0025	8208	84.64
5 SVT	0.505 [0.01,1.00]	9.154 [1,17]	13	-0.0523	1894	64.3
6 SVT	0.298 [0.03,0.55]	10 [4,18]	9	0.02	455	61.03
7 VT	0.504 [0.01,1.00]	9.357 [2,16]	14	-0.0593	5270	84.36
8 SVT	0.170 [0.01,0.48]	9.5 [7,12]	6	-0.05	460	48.64
9 SVT	0 [0,0]	0 [0,0]	1	0	279	47.72
10 VT	0.565 [0.01,1.00]	7.091 [2,13]	11	-0.0518	4739	89.34
11 SVT	0.033 [0.01,0.06]	11 [10,12]	3	-0.0267	343	45.87
12 SVT	0.326 [0.01,0.75]	11.385 [3,18]	13	-0.0077	876	59.39
13 SVT	0.084 [0.01,0.20]	16 [14,18]	5	-0.036	363	50.38
14 SVT	0.067 [0.01,0.16]	15.333 [12,18]	6	-0.01	539	52.01
15 SVT	0.498 [0.01,0.92]	13.5 [11,16]	6	0.0083	374	51.23
16 VT	0.468 [0.02,0.99]	6 [1,11]	11	-0.0064	4419	89.06
17 VT	0.490 [0.05,1.00]	10.6 [6,16]	10	-0.004	2699	84.82
18 VT	0.517 [0.04,1.00]	10.7 [6,16]	10	-0.009	2489	84.45
19 VT	0.506 [0.04,1.00]	10.6 [6,16]	10	-0.02	2812	84.87