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 (UPenn), Houssam Abbas (Oregon State)

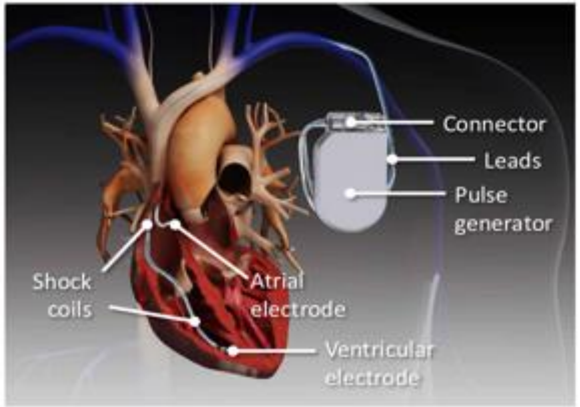
CPS-SR 2019 @ CPSWeek, Montreal, 15 April 2019

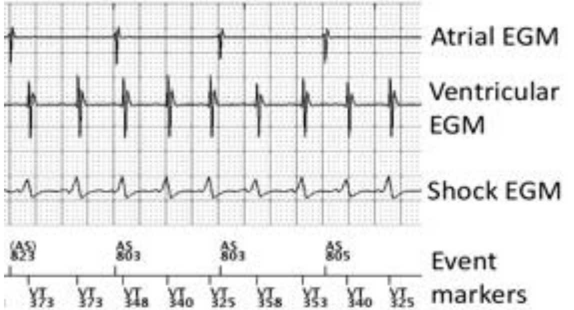
What are ICDs?

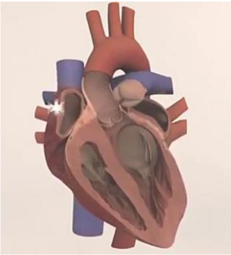
Implantable **C**ardioverter **D**efibrillators

- ○ Prevent sudden cardiac death in patients
- ○ **High-energy shocks** to terminate arrhythmia
- ○ Monitor 3 signals: atrial, ventricular, shock EGM

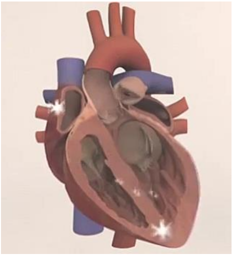
ICDs run **discrimination algorithms** to detect and treat potentially fatal arrhythmias from EGM signals







Normal sinus rhythm

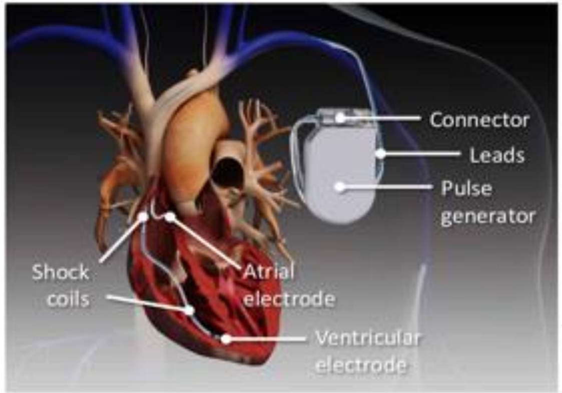


Ventricular fibrillation

ICD communication

In-clinic settings

Patient



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

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)

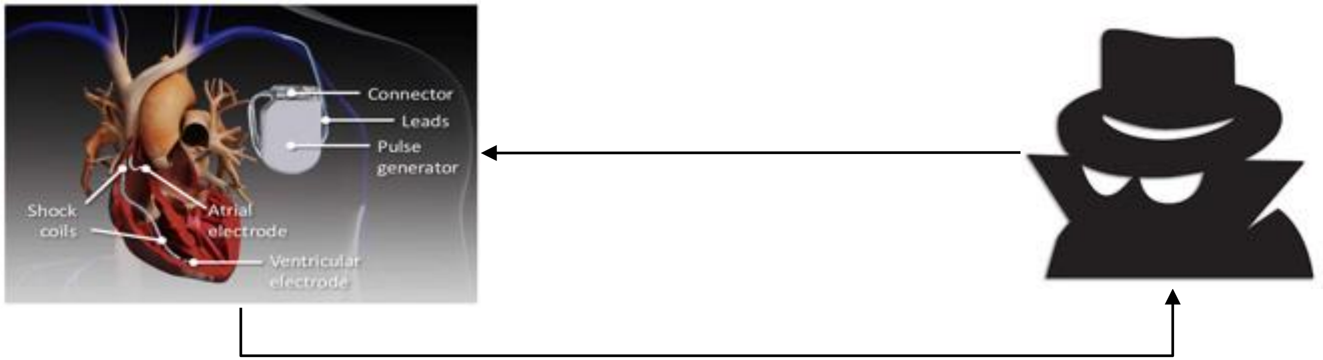
Medtronic MyCareLink Smart™
The reader (left) interrogates the ICD and sends medical data to smartphone app via Bluetooth

Security Concerns

- 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)

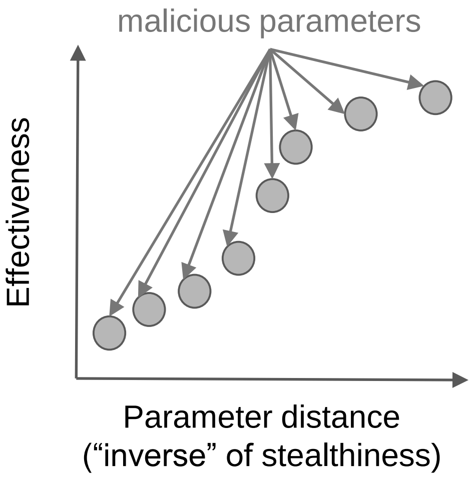
Aim of this study

- ICD unauthorized access is possible exploiting unsecure wireless link
- **Can one reprogram an ICD to affect therapy without being detected?**
- We present a systematic method to do so



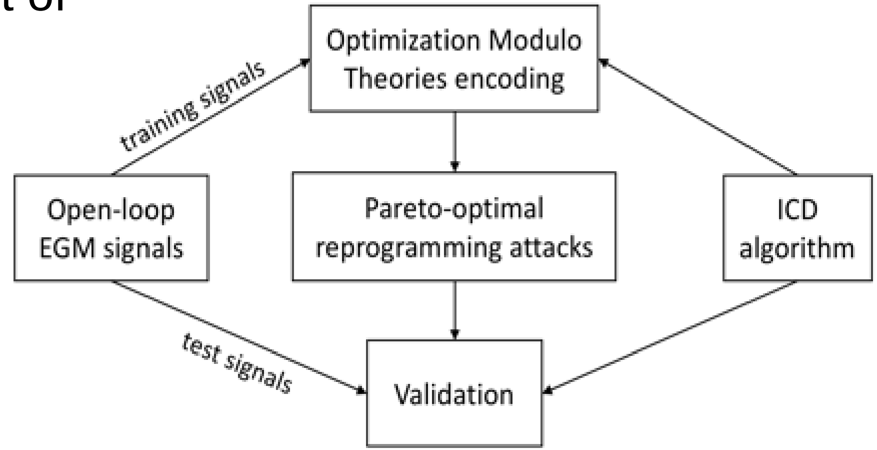
Synthesizing Stealthy Attacks on ICDs

- 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

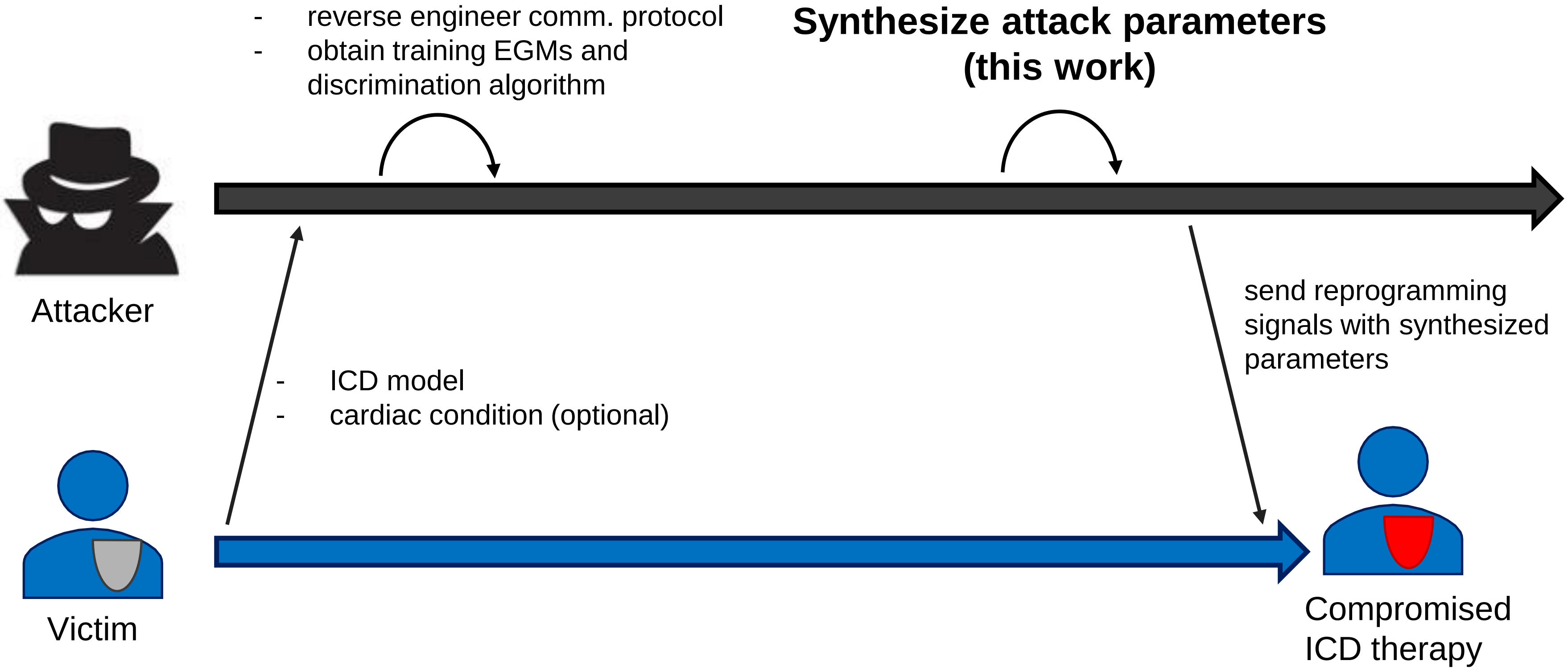


Methodology Overview

- 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 – Timeframe



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

Synthesize attack parameters (this work)

Attacker

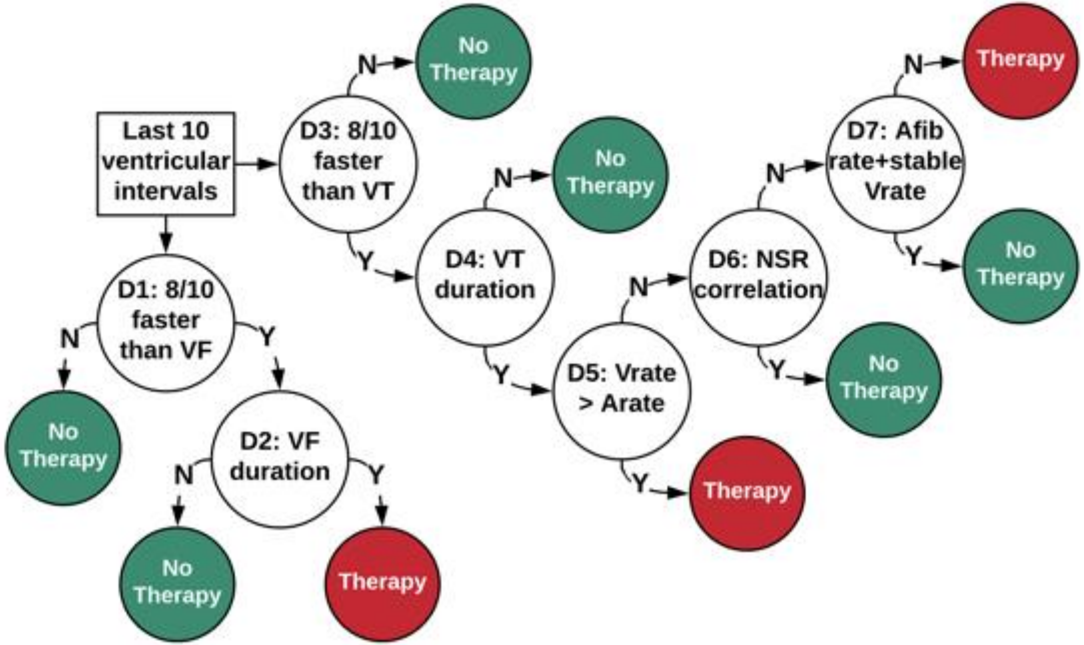
- ICD model
- cardiac condition (optional)

send reprogramming signals with synthesized parameters

Victim

Compromised ICD therapy

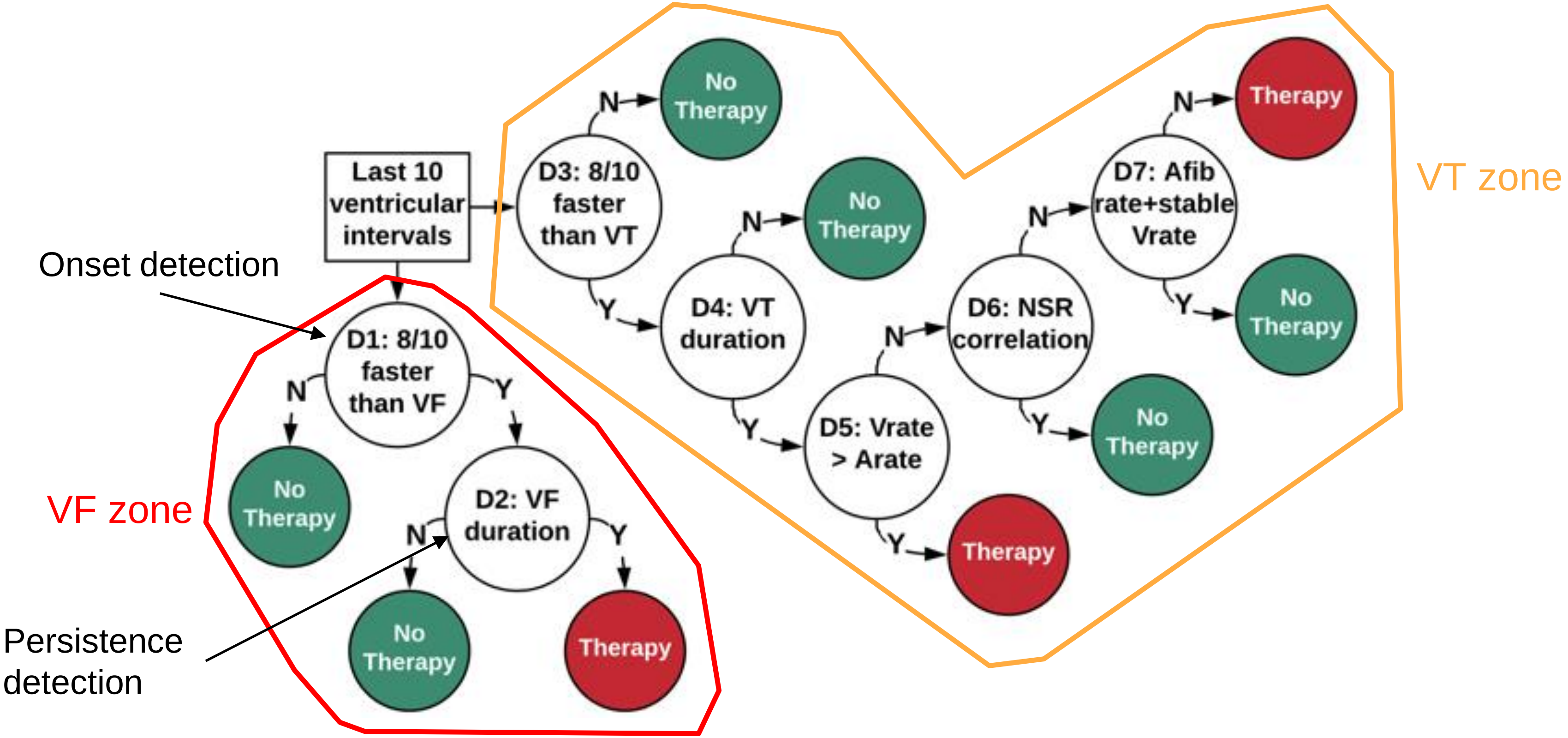
Boston Scientific ICD



B.Sc. 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 – episode detection

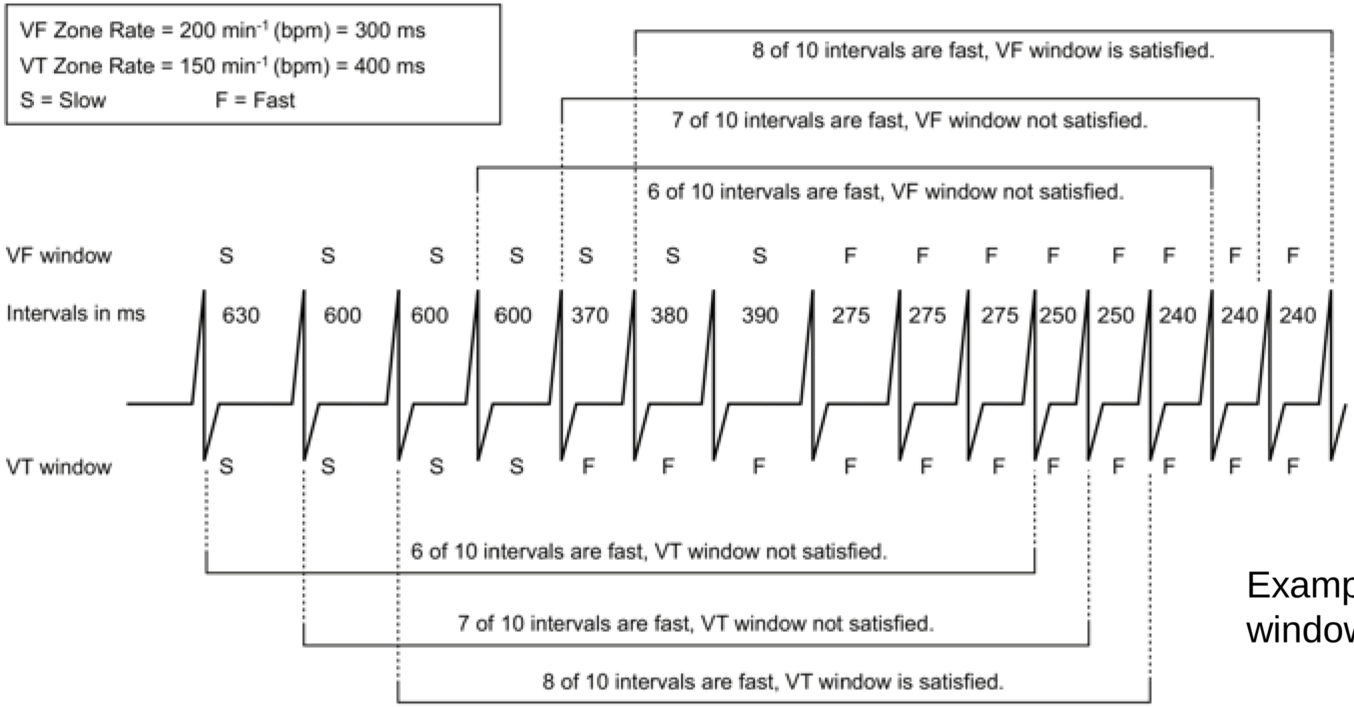
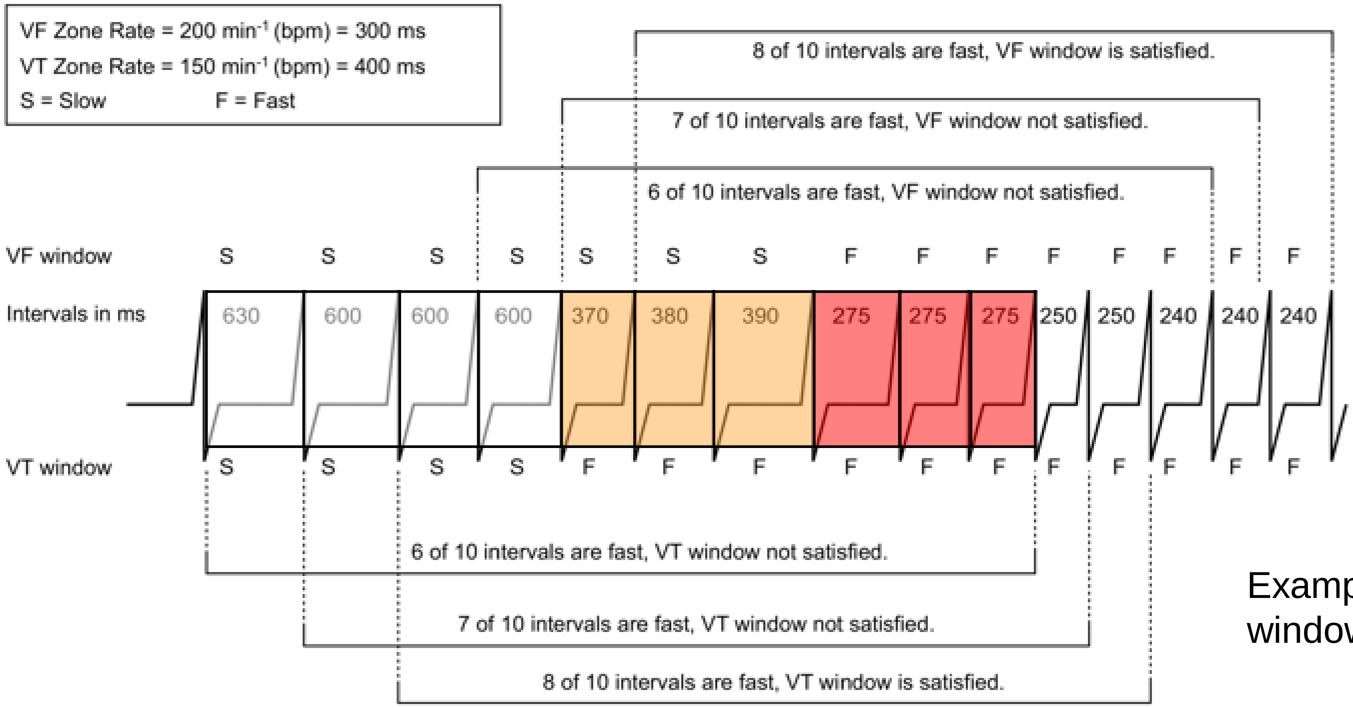

VF Zone Rate = 200 min^{-1} (bpm) = 300 ms
VT Zone Rate = 150 min^{-1} (bpm) = 400 ms
S = Slow F = Fast
8 of 10 intervals are fast, VF window is satisfied.
7 of 10 intervals are fast, VF window not satisfied.
6 of 10 intervals are fast, VF window not satisfied.
VF window
S S S S S S S F F F F F F F F
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 – episode detection



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 – episode detection

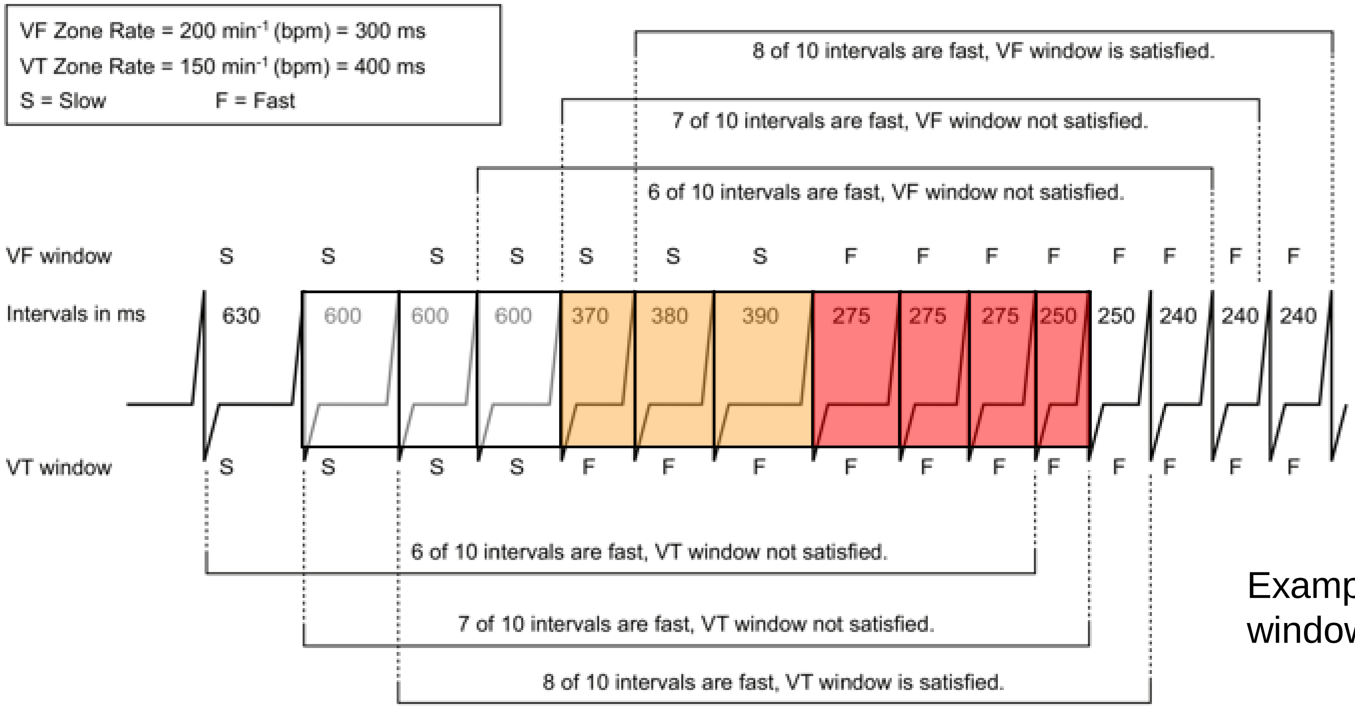

VF Zone Rate = 200 min⁻¹ (bpm) = 300 ms
VT Zone Rate = 150 min⁻¹ (bpm) = 400 ms
S = Slow
F = Fast
8 of 10 intervals are fast, VF window is satisfied.
7 of 10 intervals are fast, VF window not satisfied.
6 of 10 intervals are fast, VF window not satisfied.
VF window
S S S S S S S F F F F F F F F
Intervals 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.

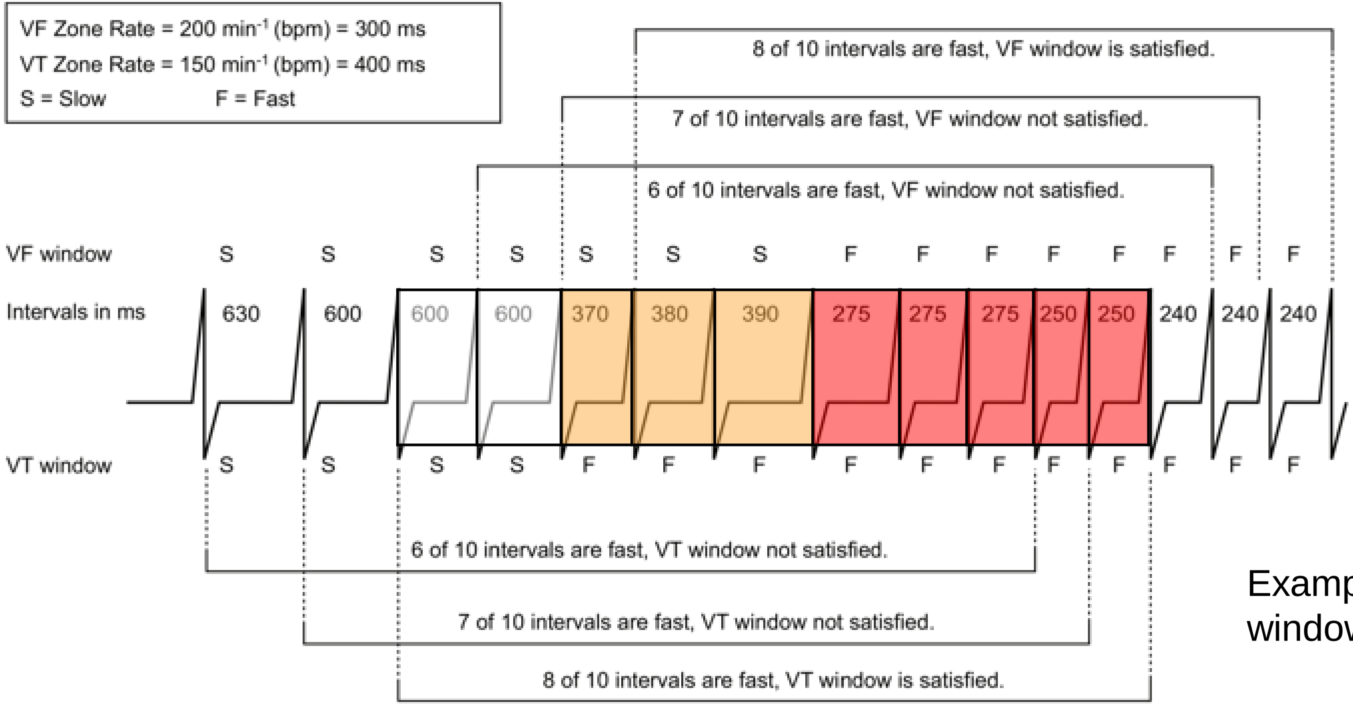
Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

🟧 Faster than VT

🟥 Faster than VF

Example of detection windows (BSc ICD manual)

Boston Scientific ICD – episode detection



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.

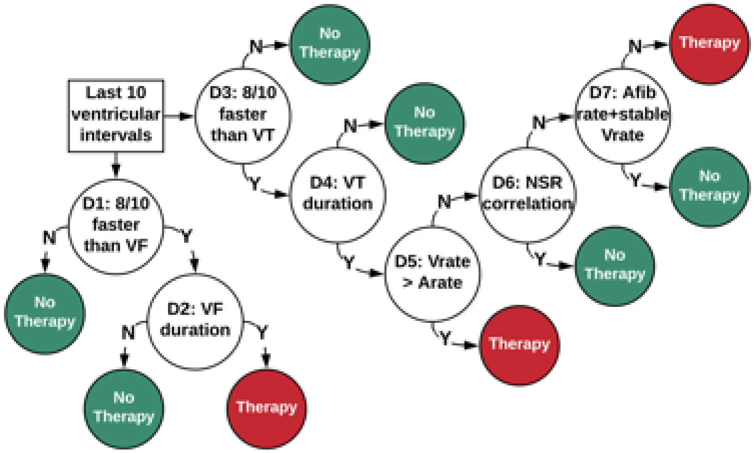
Figure 2–4. Interaction of ventricular detection windows, 2-zone configuration

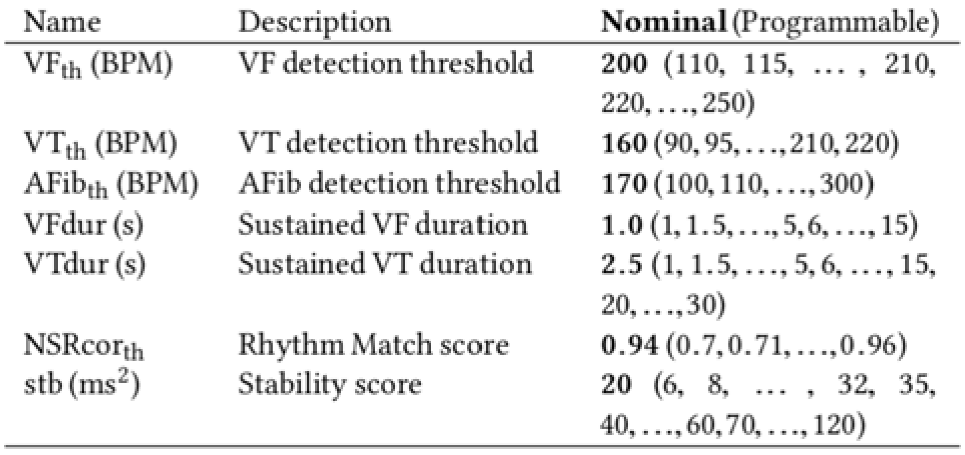
Faster than VT

Faster than VF

Example of detection windows (BSc ICD manual)

Boston Scientific ICD – parameters

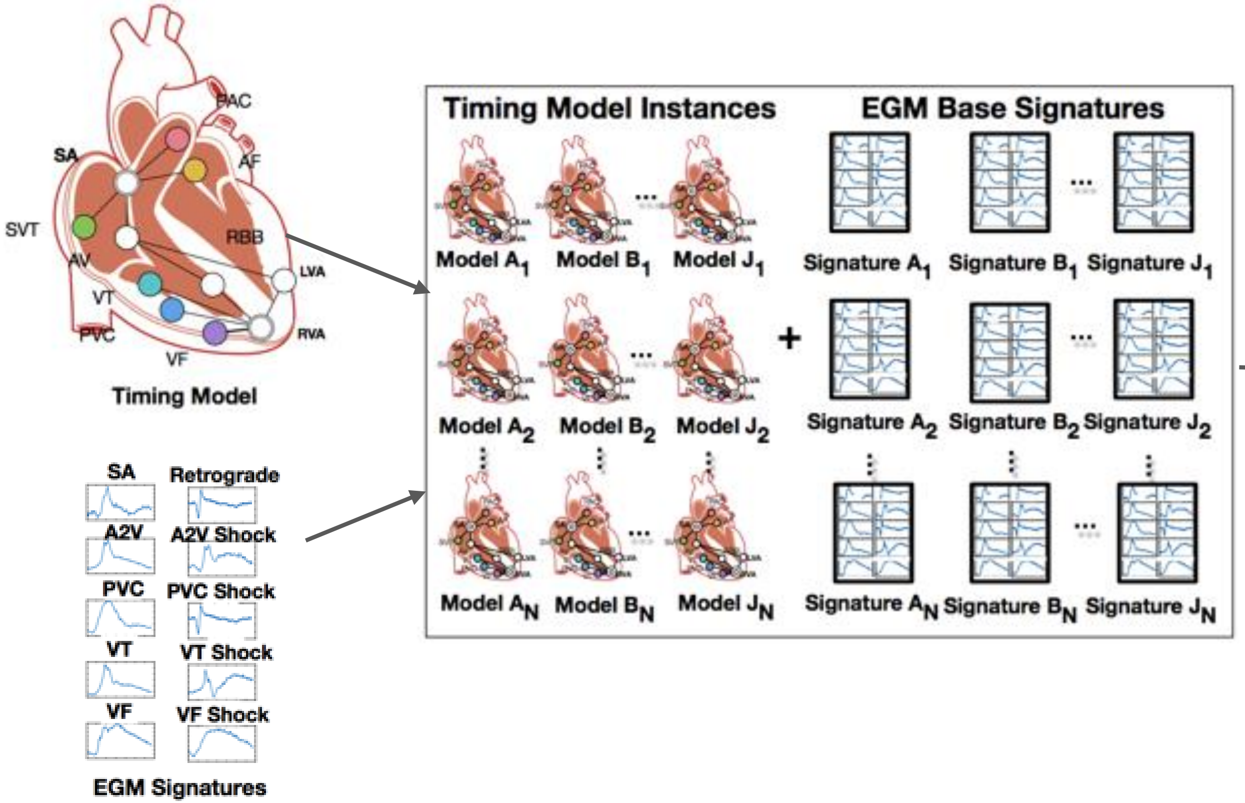


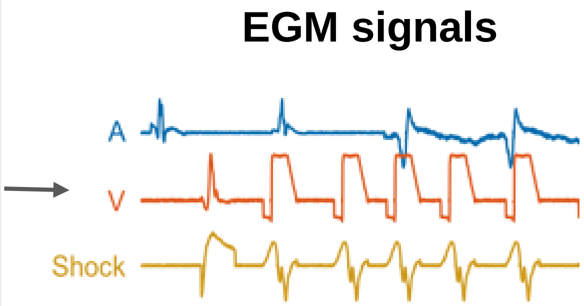


Name	Description	**Nominal** (Programmable)
VF_{th} (BPM)	VF detection threshold	**200** (110, 115, ... , 210, 220, . . ., 250)
VT_{th} (BPM)	VT detection threshold	**160** (90, 95, . . .,210, 220)
$AFib_{th}$ (BPM)	AFib detection threshold	**170** (100, 110, . . ., 300)
VFdur (s)	Sustained VF duration	**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]





19 different heart conditions:

- Positive (require therapy)
- Negative (no therapy)

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**

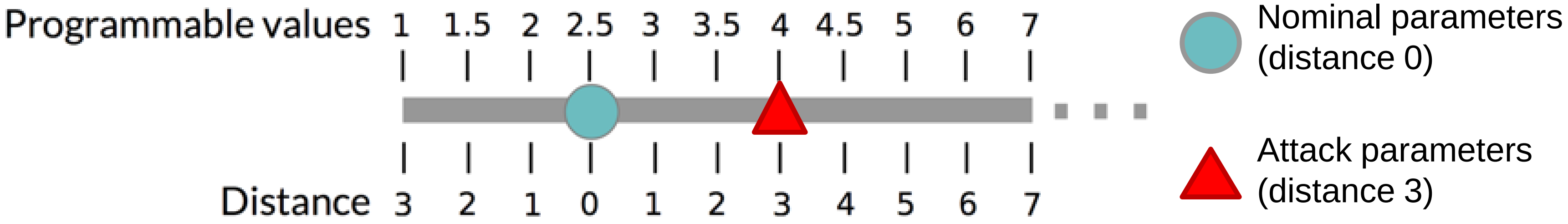
True iff therapy is given at any point in **s** under nominal parameters **p***

Attack stealthiness

“An attack is stealthy when the deviation from the nominal parameters is small”

Deviation = number of programmable values separating nominal and attack parameters (max separation over all parameters)

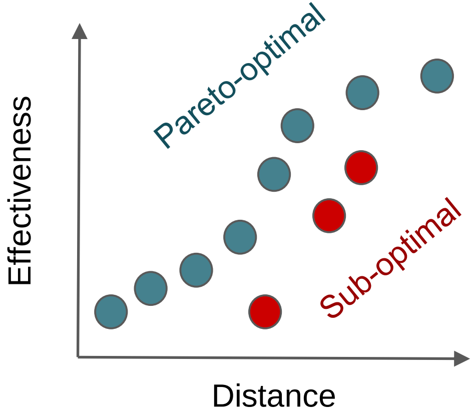
Example: parameter VT duration (s)



Synthesis of optimal stealthy attacks

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

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



<u>Challenging optimization problem</u>

- nonlinear, non-convex, combinatorial, constrained by ICD algorithm

Solution via optimization modulo theories (OMT)

- SMT (SAT + theories) is well-suited to solve combinatorial problems
  [De moura and Bjørner, 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 and nonlinear operations 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 models (among all SAT assignments) that optimize some objectives

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?

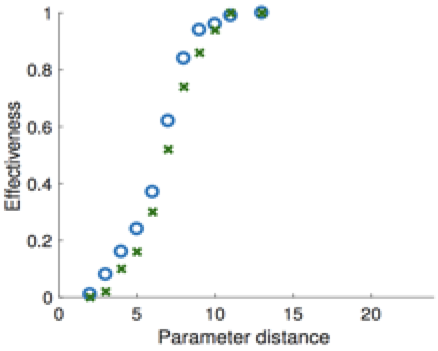
In VT duration?

Time spent in VFd

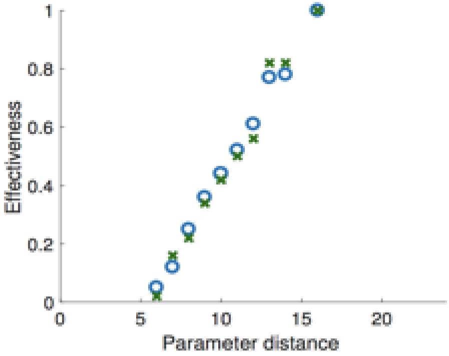
Time spent in VTd

Evaluation, condition-specific attacks

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



Condition 10
(positive)



Condition 17
(positive)

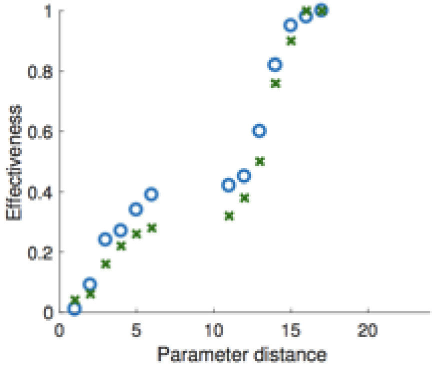
○ Training signals ✖ Validation signals

- Attacks on “positive” conditions are all very effective
- But not all equally stealthy (see left)

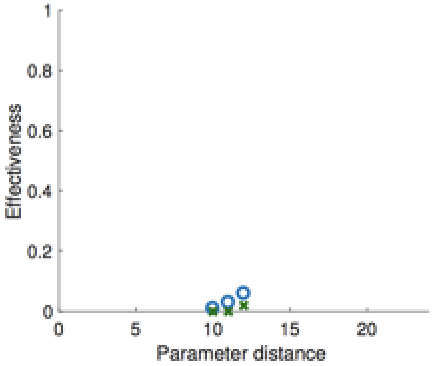
Common attack strategy:

- Increase VT and VF detection thresholds to reduce detection rate
- Increase VF and VT durations to reduce probability that episode is marked sustained

Evaluation, condition-specific attacks



Condition 5
(negative)



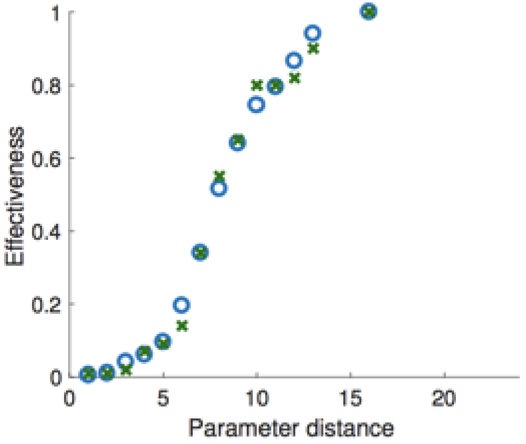
Condition 11
(negative)

○ Training signals ✖ Validation signals

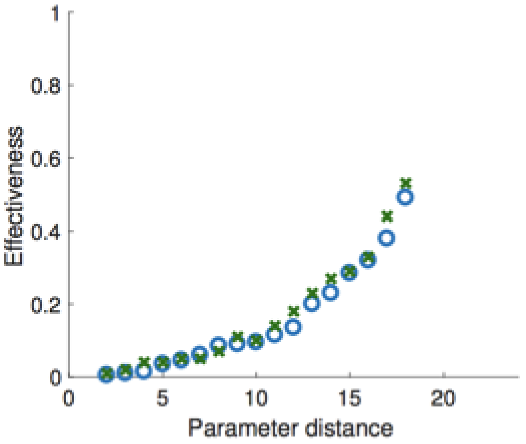
- Attacks on negative conditions are not all equally effective
- Because, under normal HR, VT and VF must be reprogrammed to very low values to classify it as fast HR
- *Common attack strategy*: keep VF/VT thresholds and duration to a minimum

Evaluation, condition-agnostic attacks

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



positive conditions

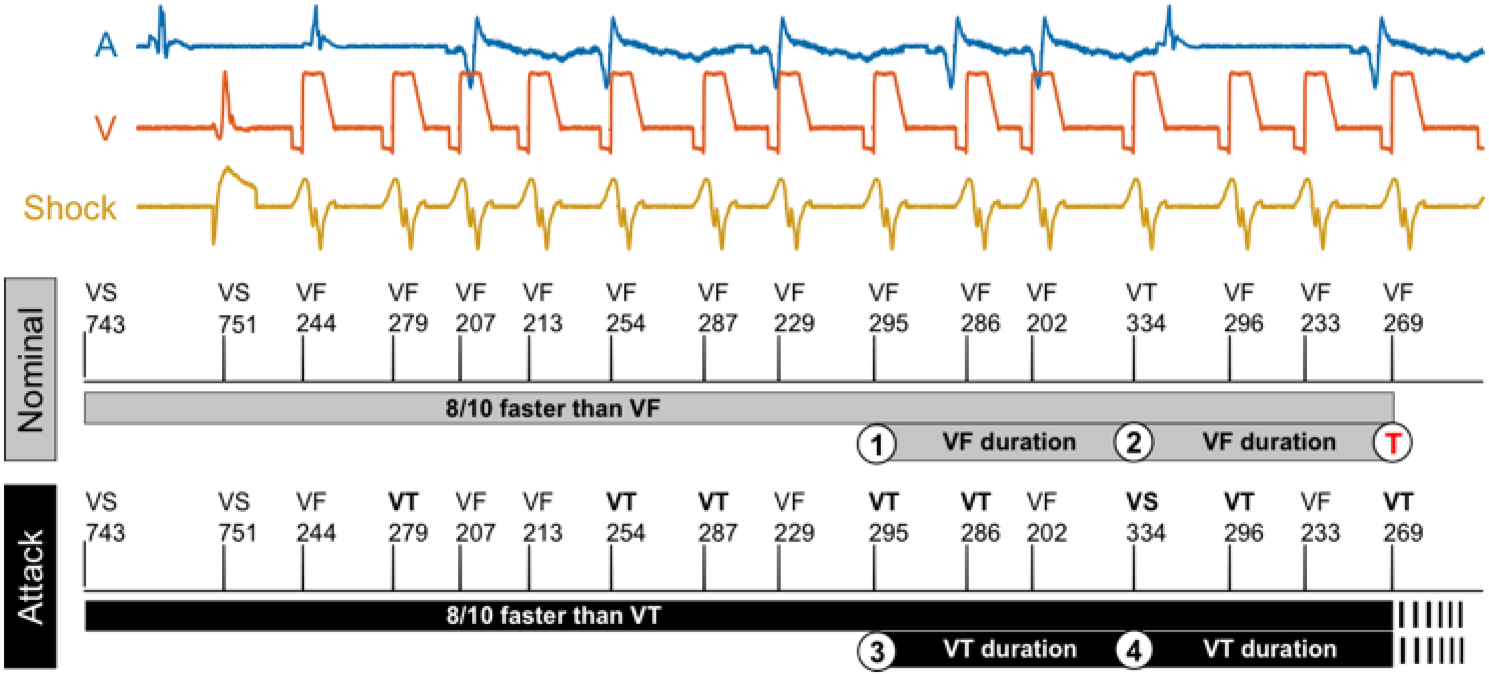


negative conditions

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

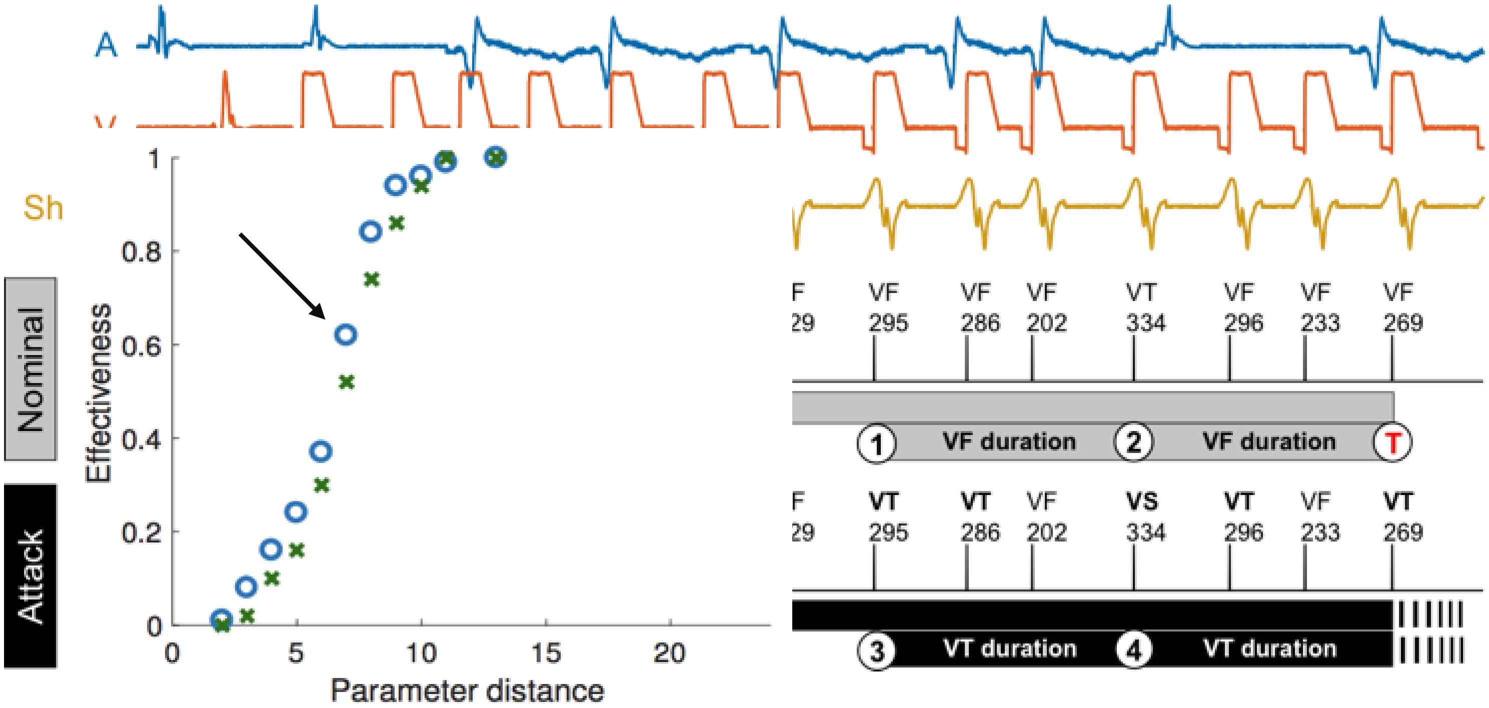


EGM extract from condition 10 signals

Evaluation, condition-specific attacks

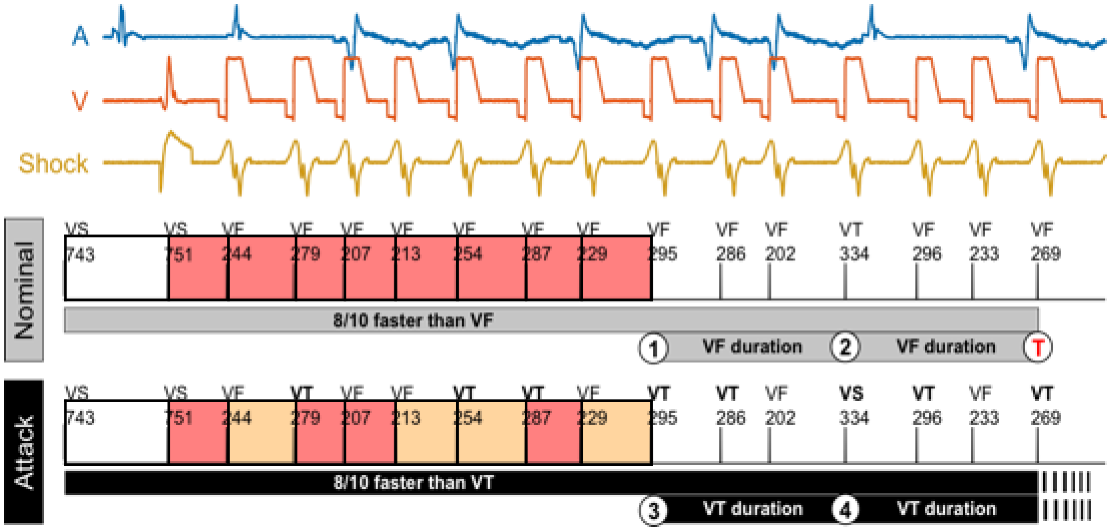
VF_th = 200 BPM
VT_th = 160 BPM
VFdur = 1 s
VTdur = 2.5 s

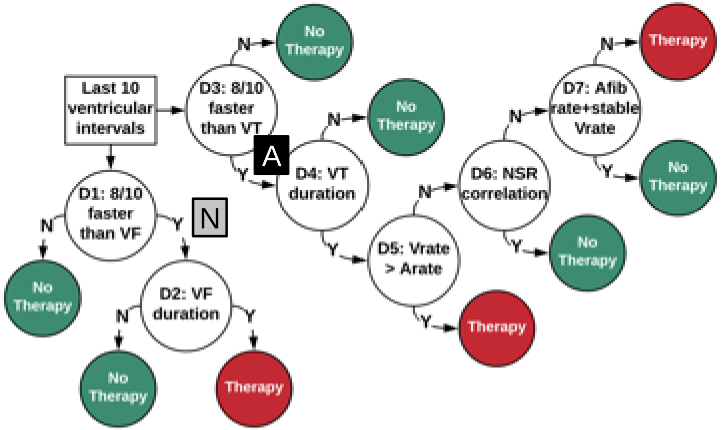
VF_th = 240 BPM
VT_th = 185 BPM
VFdur = 4 s
VTdur = 7 s



EGM extract from condition 10 signals

Evaluation, condition-specific attacks

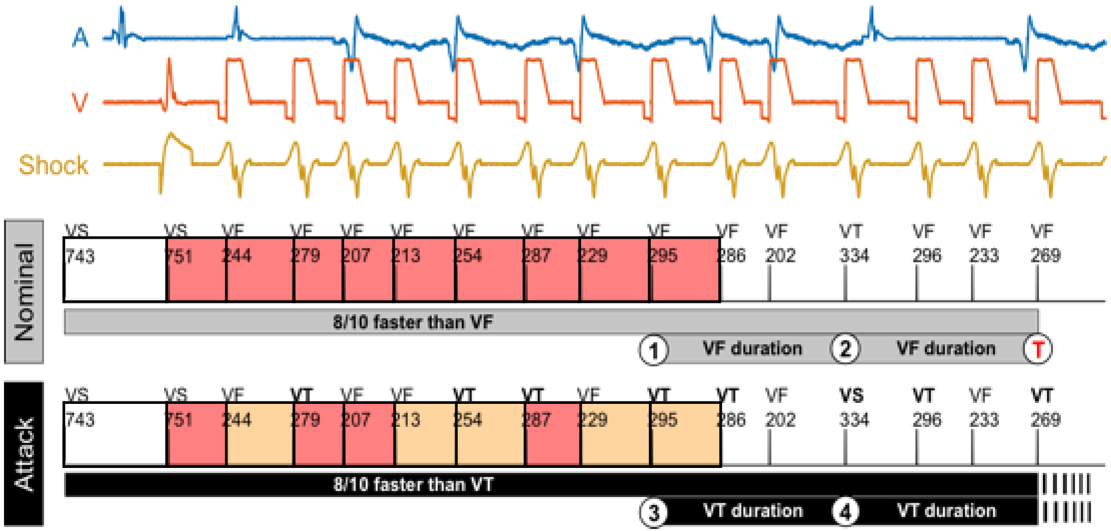


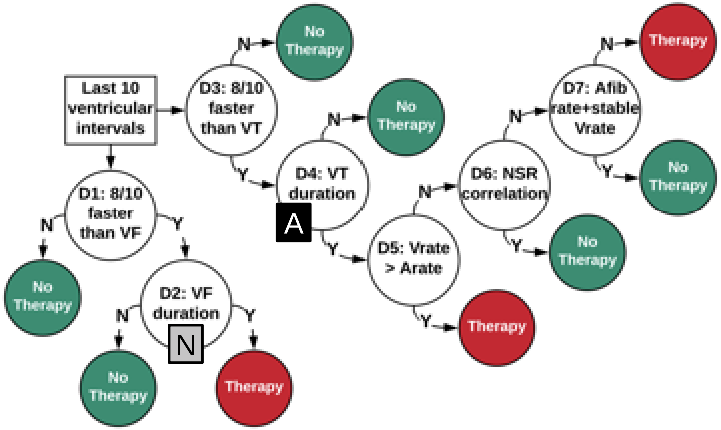


Faster than VT

Faster than VF

Evaluation, condition-specific attacks

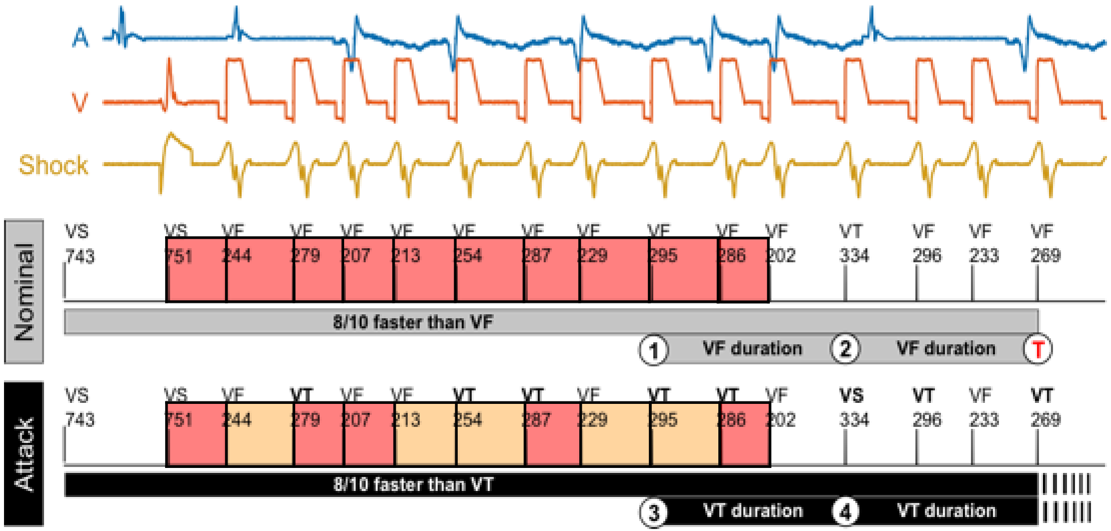


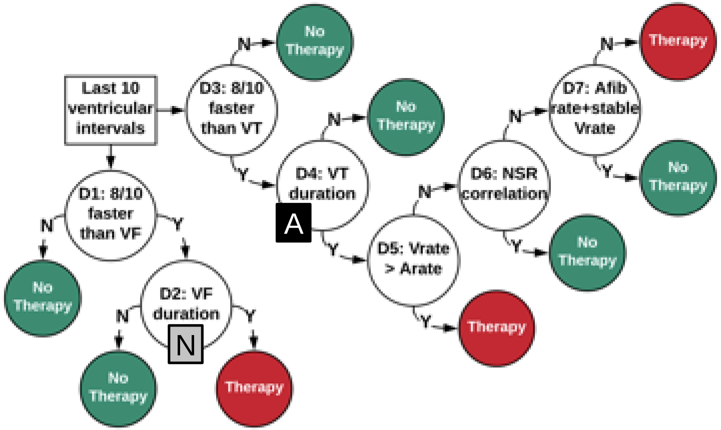


Faster than VT

Faster than VF

Evaluation, condition-specific attacks

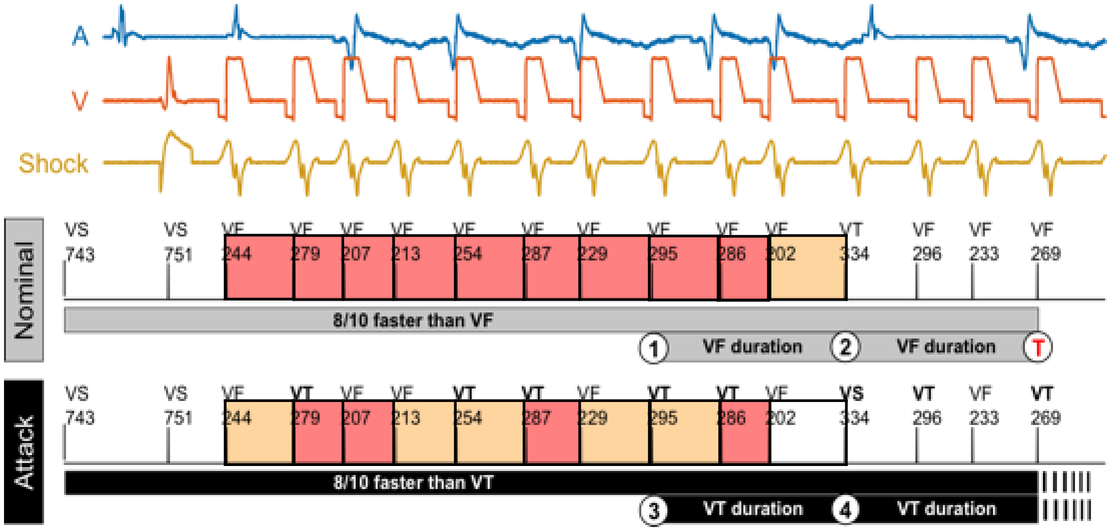


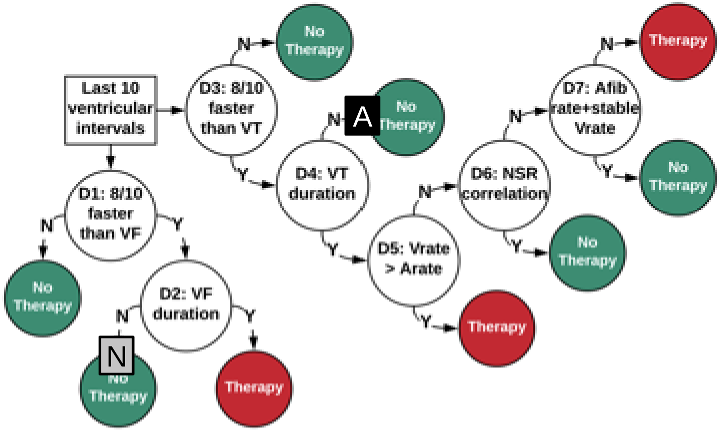


🟧 Faster than VT

🟥 Faster than VF

Evaluation, condition-specific attacks

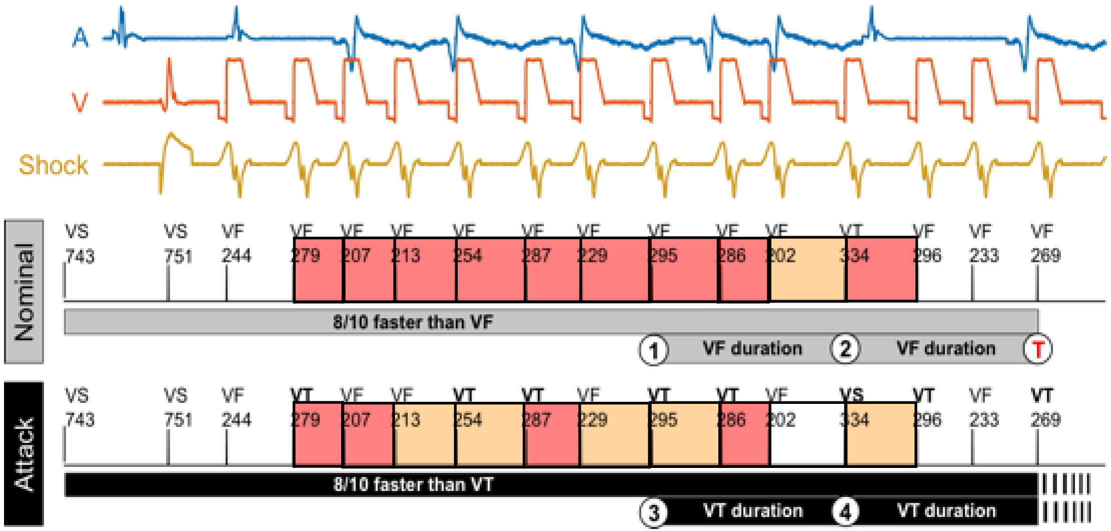


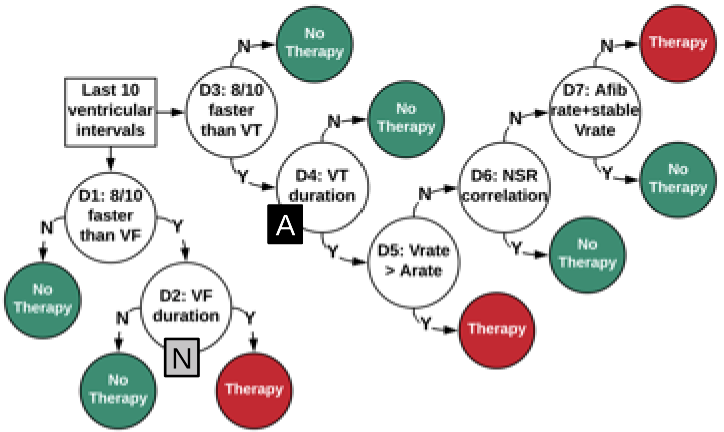


🟧 Faster than VT

🟥 Faster than VF

Evaluation, condition-specific attacks

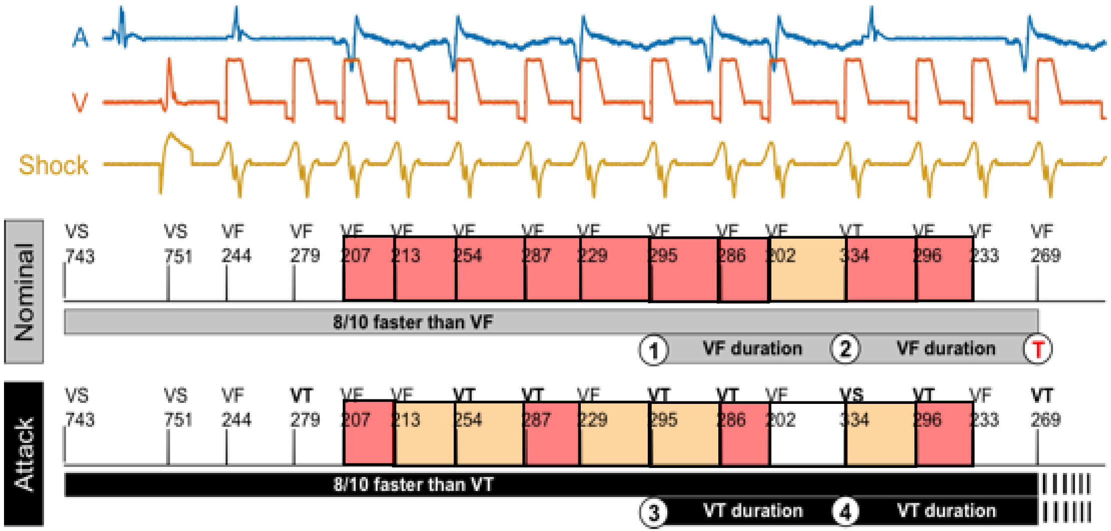


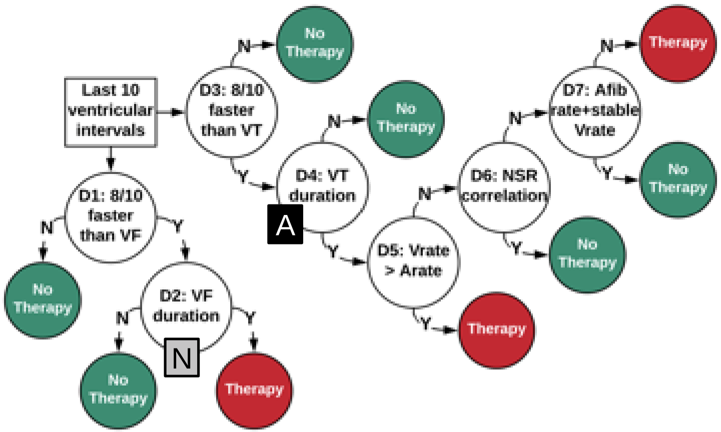


Faster than VT

Faster than VF

Evaluation, condition-specific attacks

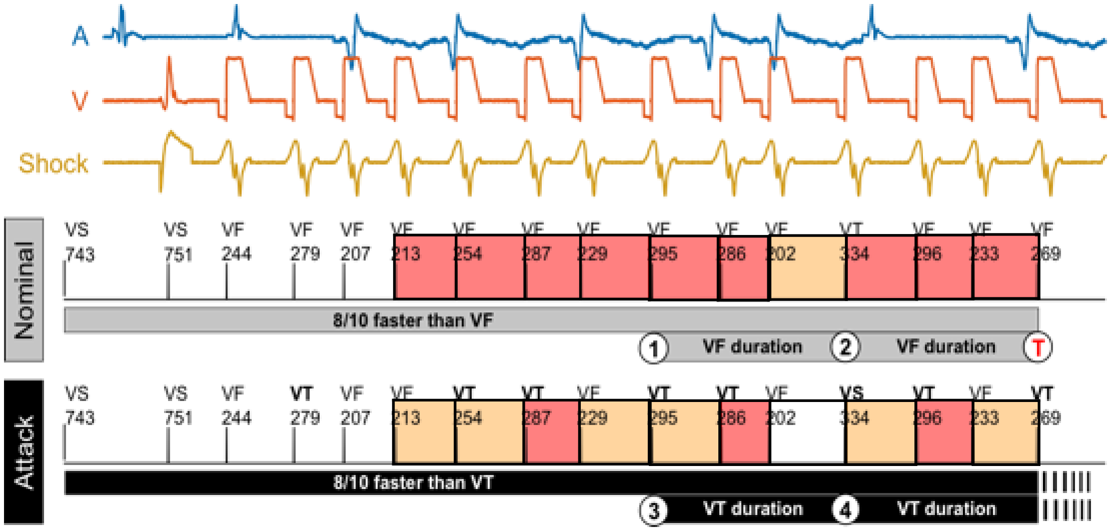


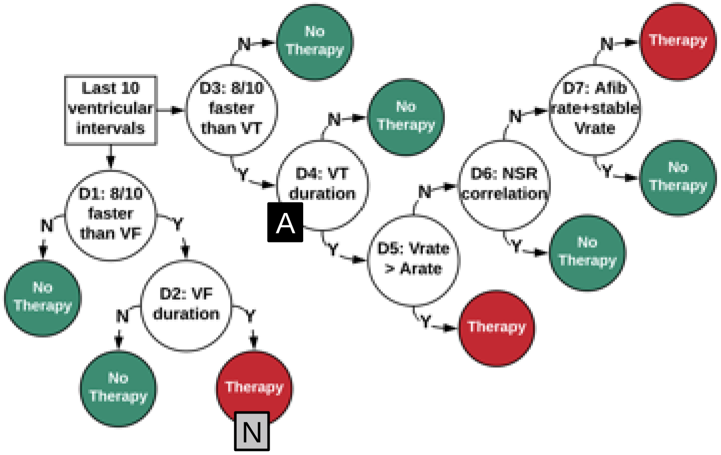


Faster than VT

Faster than VF

Evaluation, condition-specific attacks





Therapy prevented by attack

Faster than VT

Faster than VF

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