Beauty and the Burst

Remote Identification of Encrypted Video Streams

Roei Schuster
Cornell Tech,
Tel Aviv University

Vitaly Shmatikov
Cornell Tech

Eran Tromer
Columbia University,
Tel Aviv University

Video traffic is interesting

Watching What You See on the Web

New Gear Lets ISPs Track Users and Sell Targeted Ads; More Players, Privacy Fears

How ISPs can sell your Web history—and how to stop them

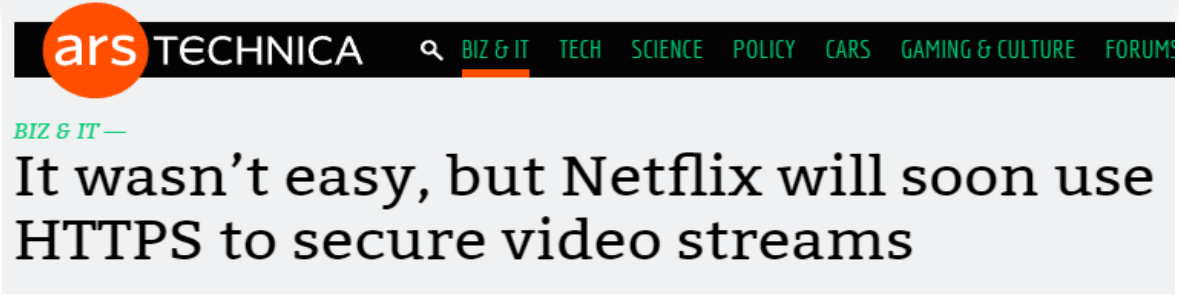
VANITY FAIR

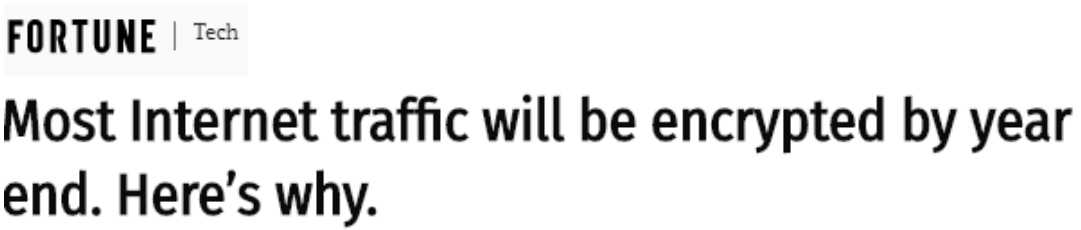
Company That Kept Trying to Guess Netflix's Ratings Finally Gives Up

Symphony Advanced Media is calling it quits.

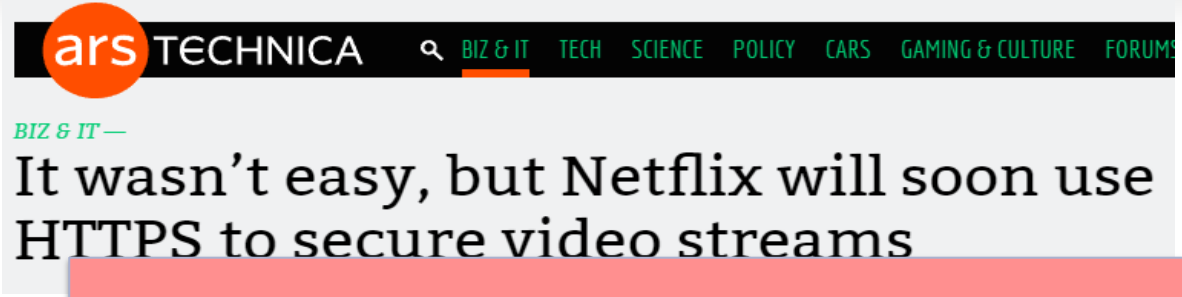
Nielsen shines a light on Netflix viewership

Video traffic is encrypted

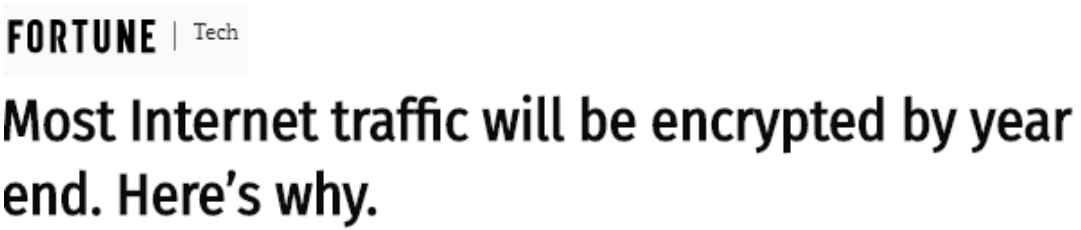




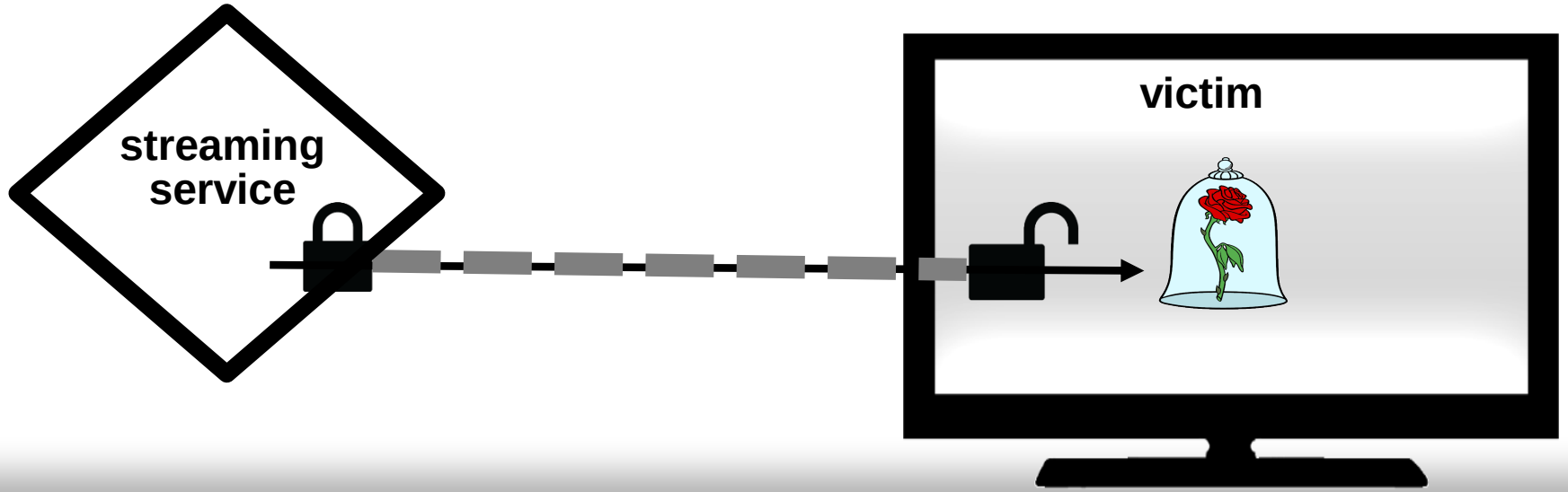
Video traffic is encrypted



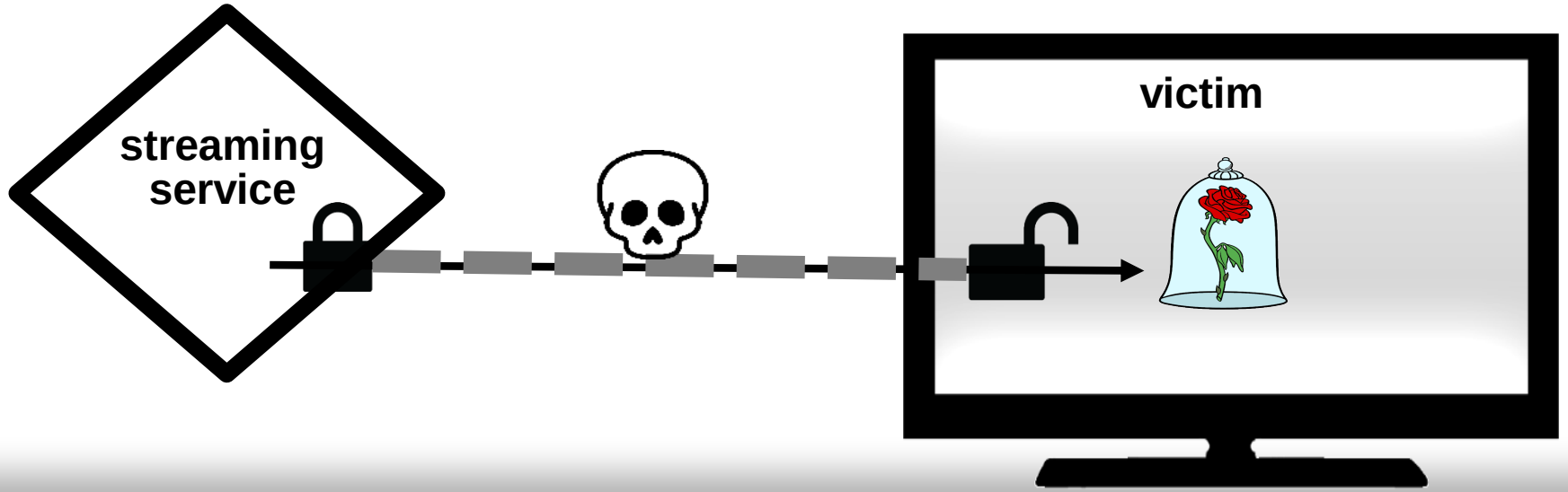
What can still be learned?



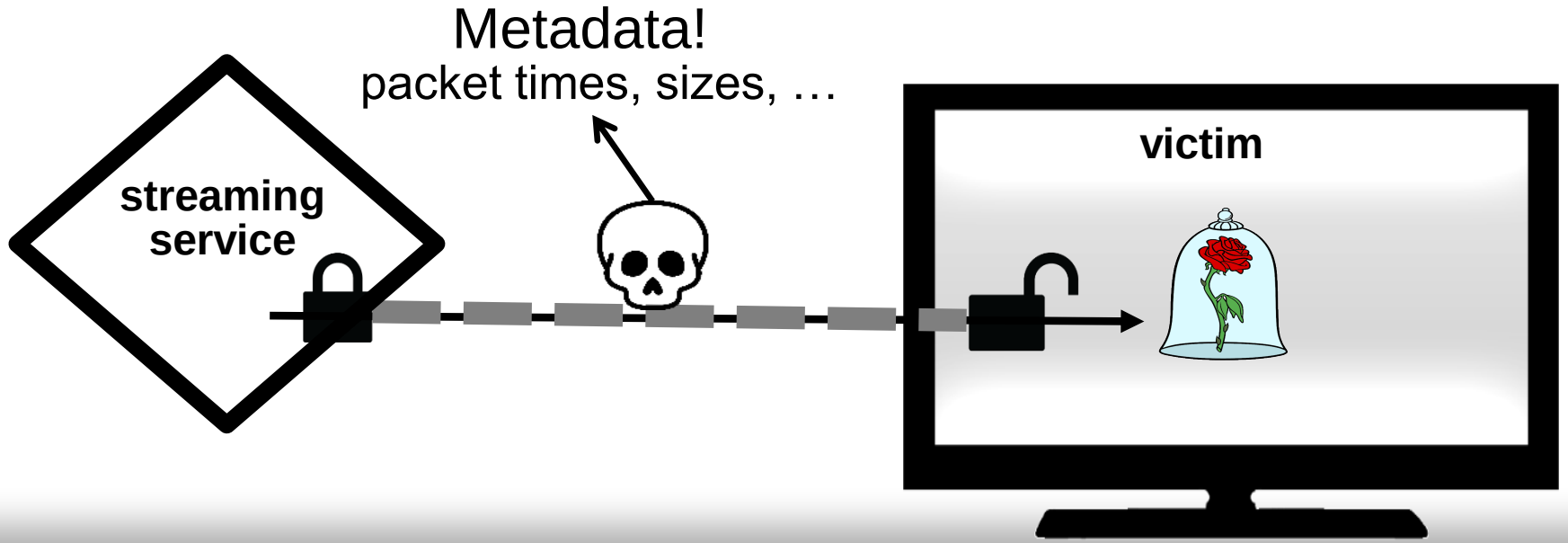
Traffic analysis for video identification



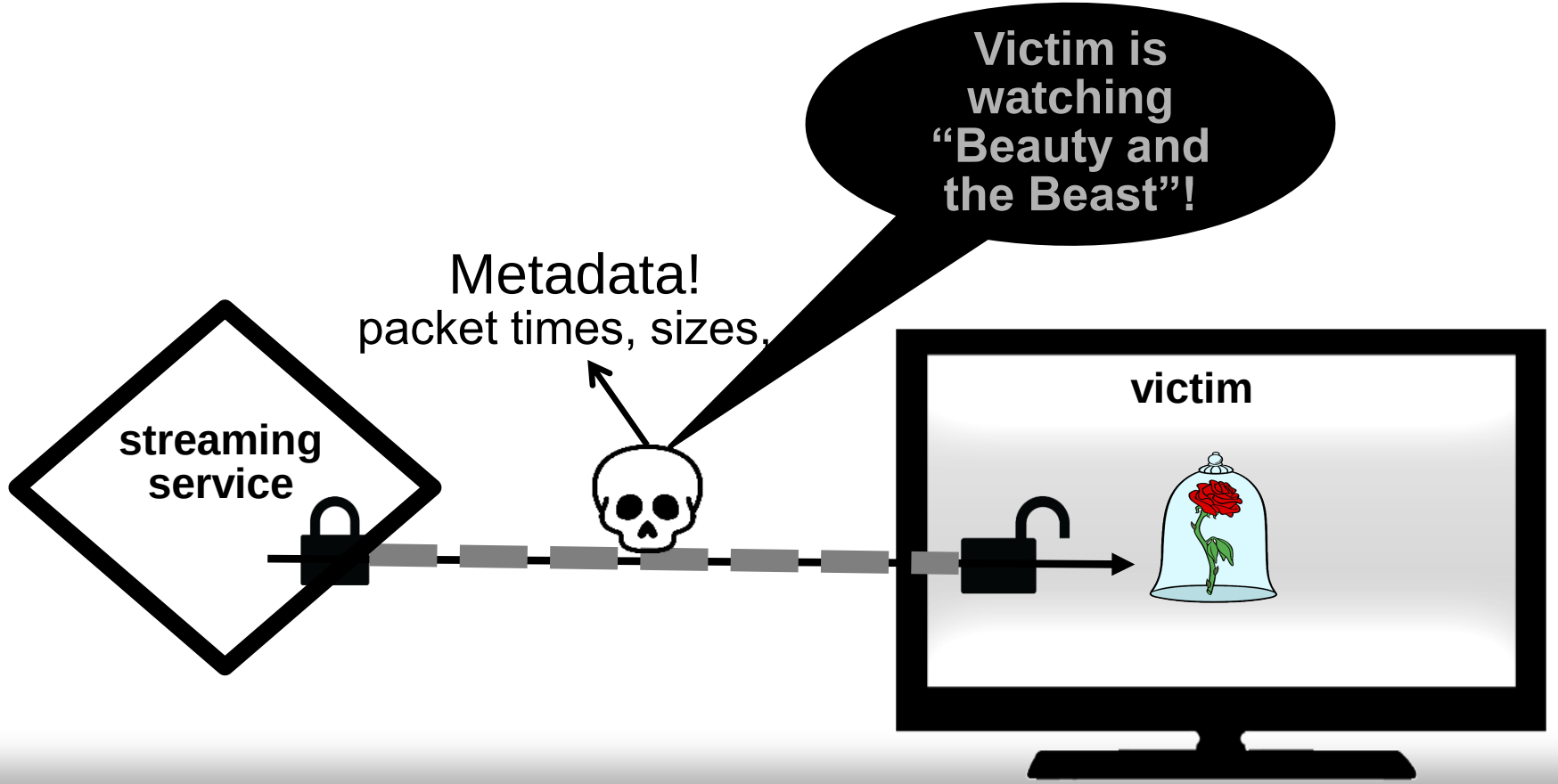
Traffic analysis for video identification



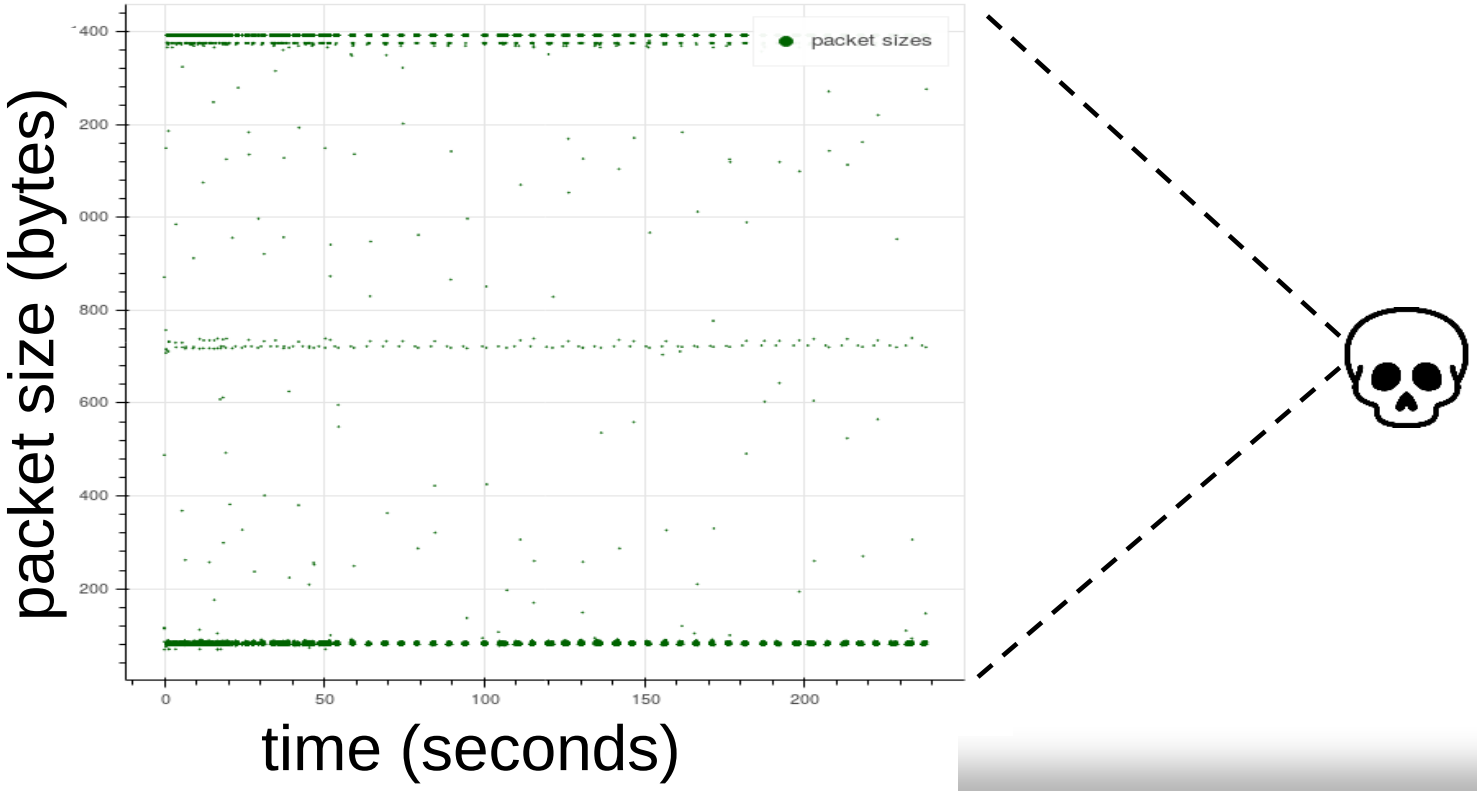
Traffic analysis for video identification



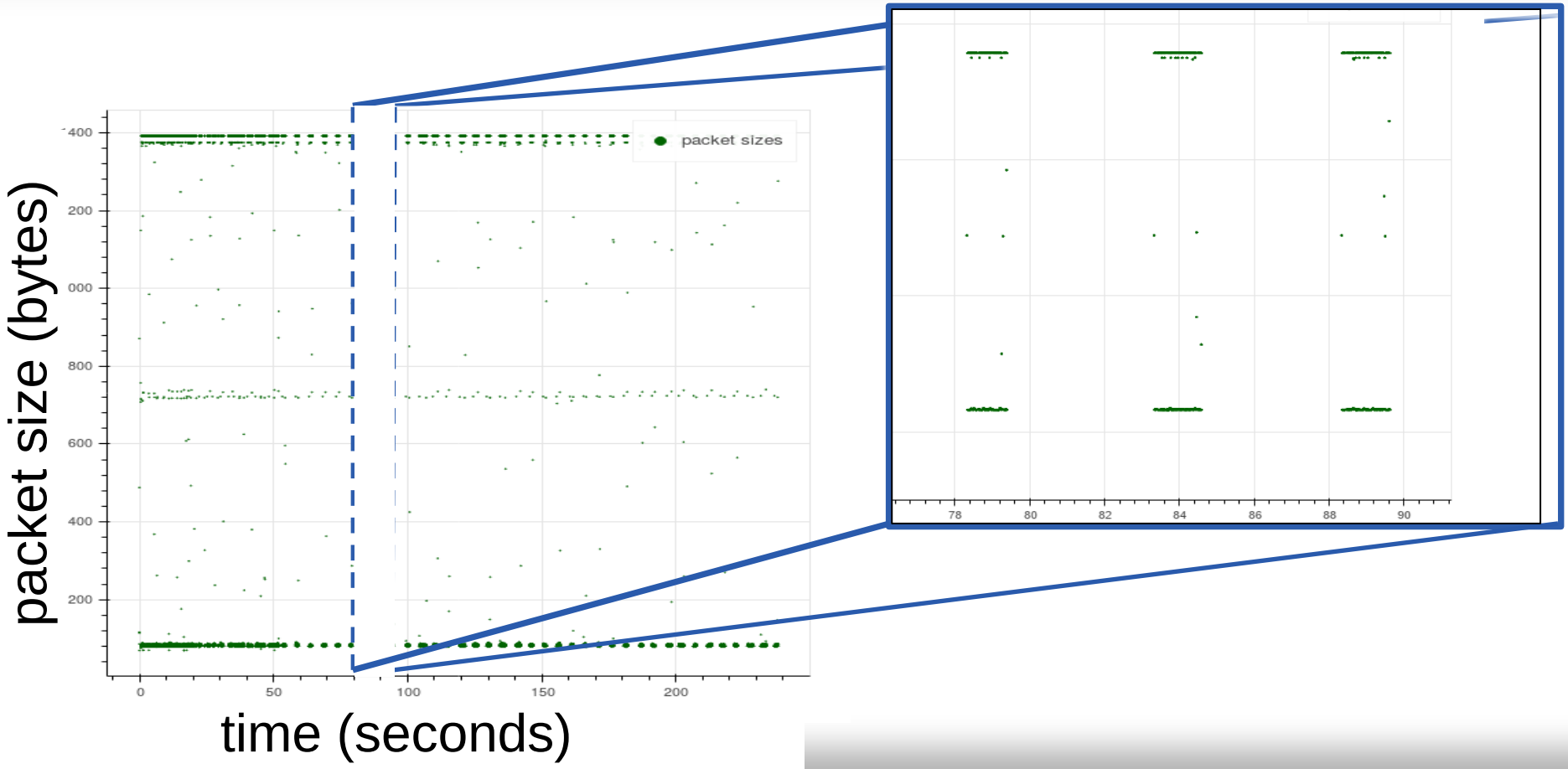
Traffic analysis for video identification



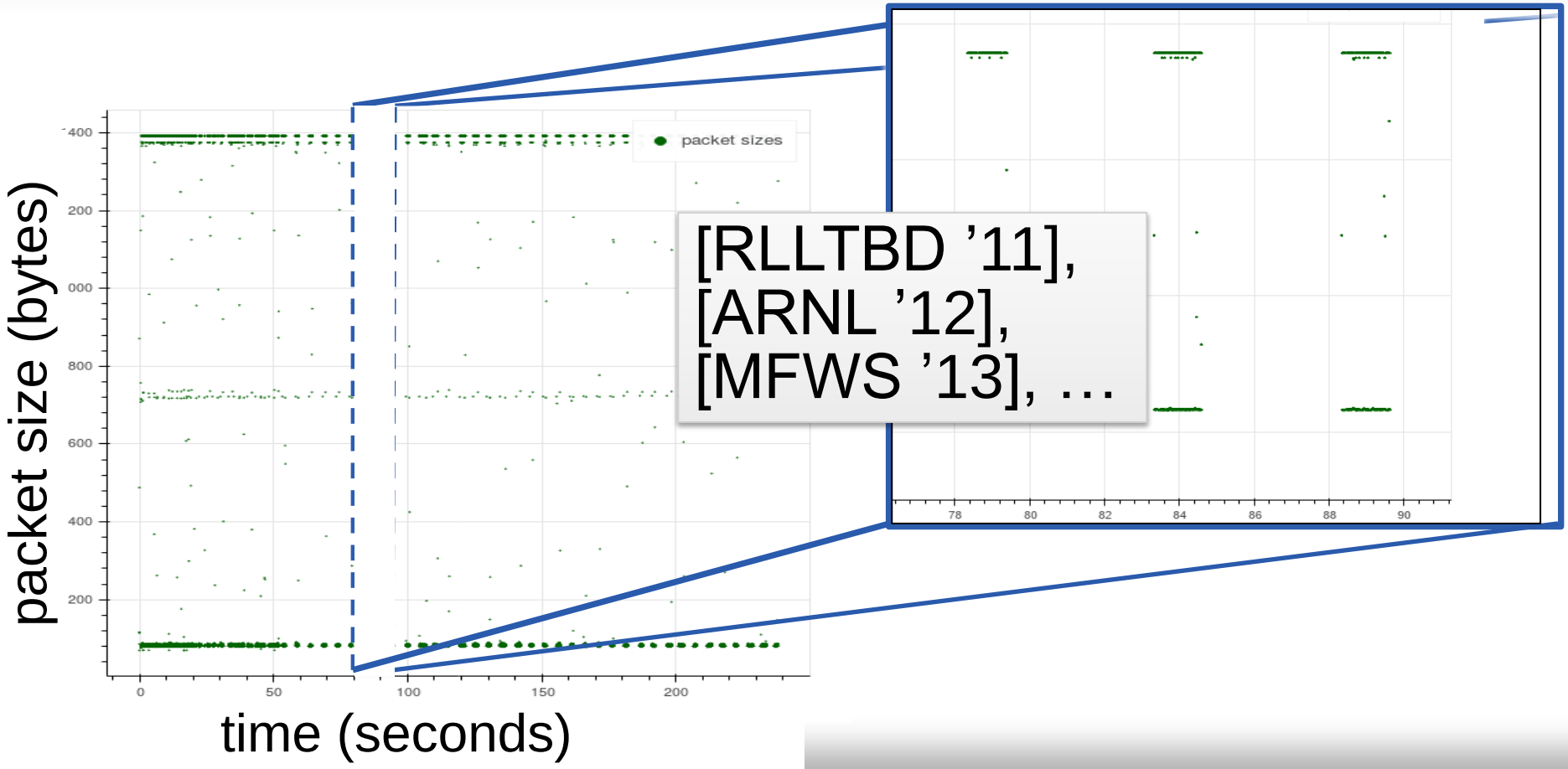
Initial buffering, then “on”/“off” bursts



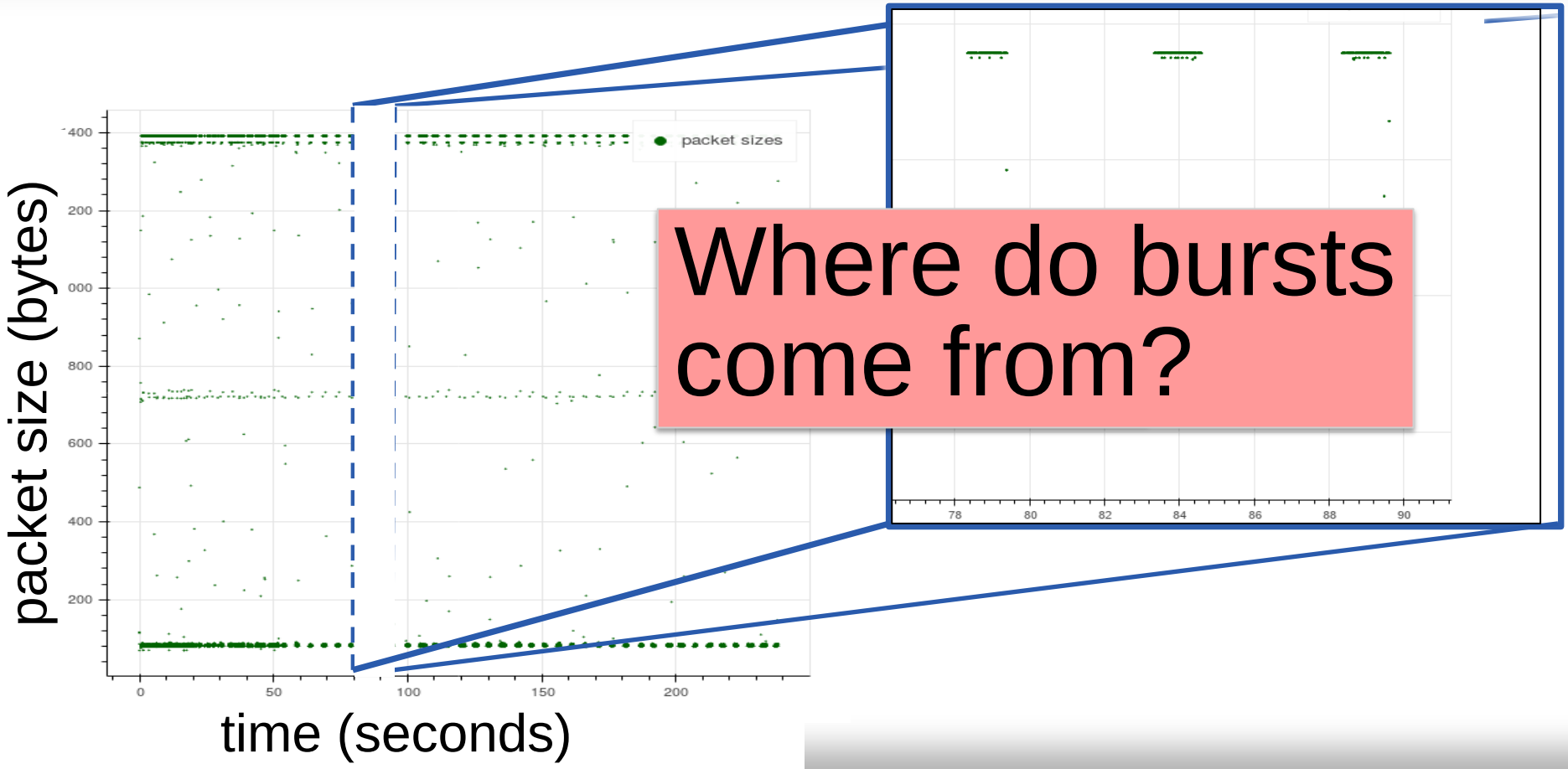
Initial buffering, then “on”/“off” bursts



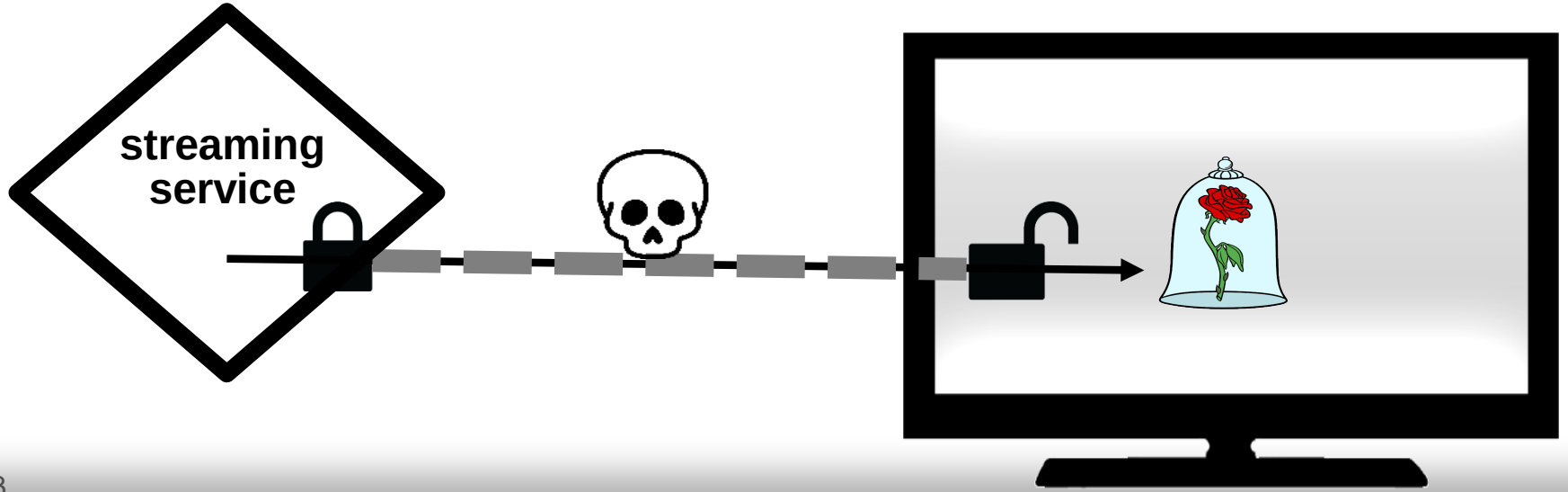
Initial buffering, then “on”/“off” bursts



Initial buffering, then “on”/“off” bursts

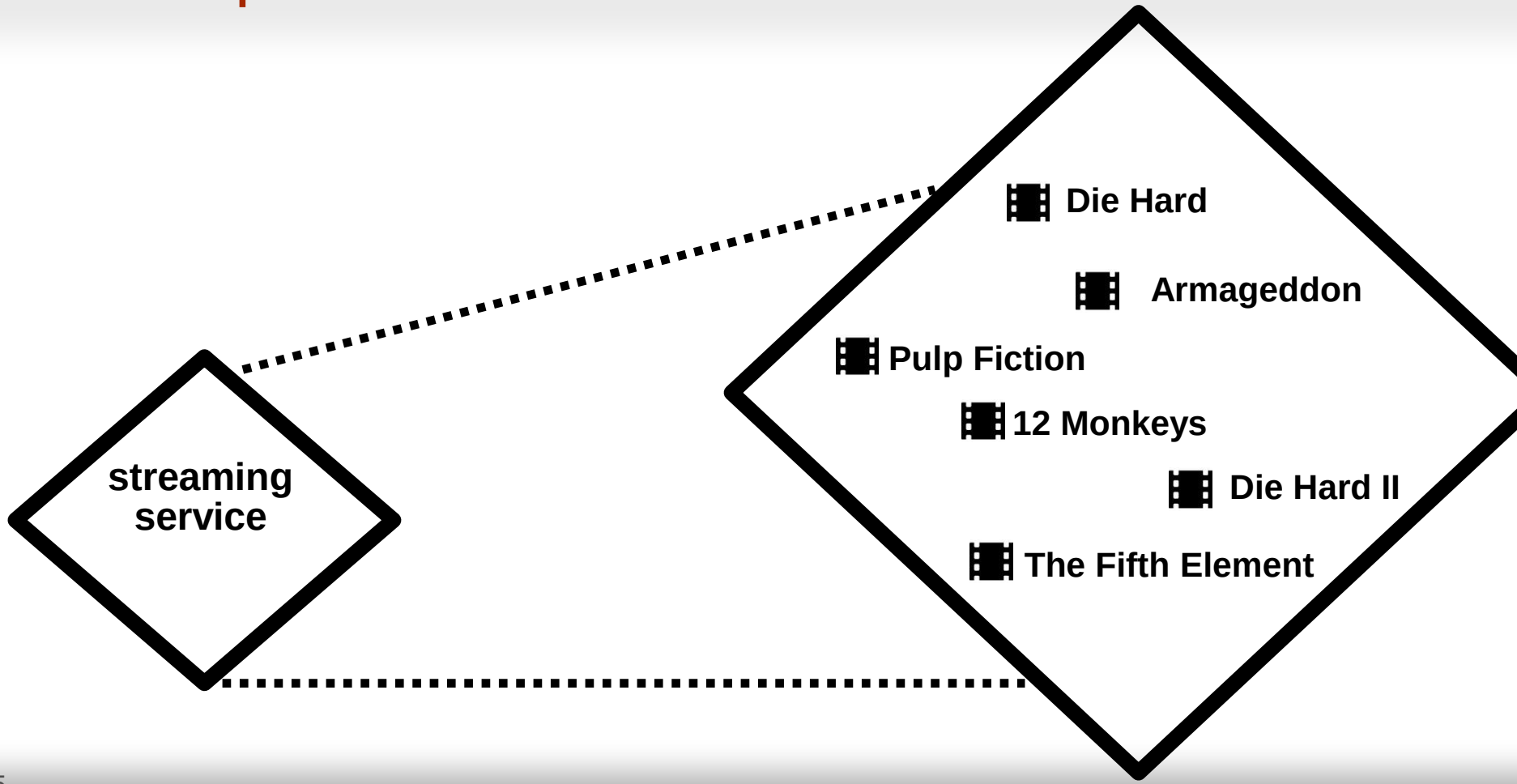


Video representation on server



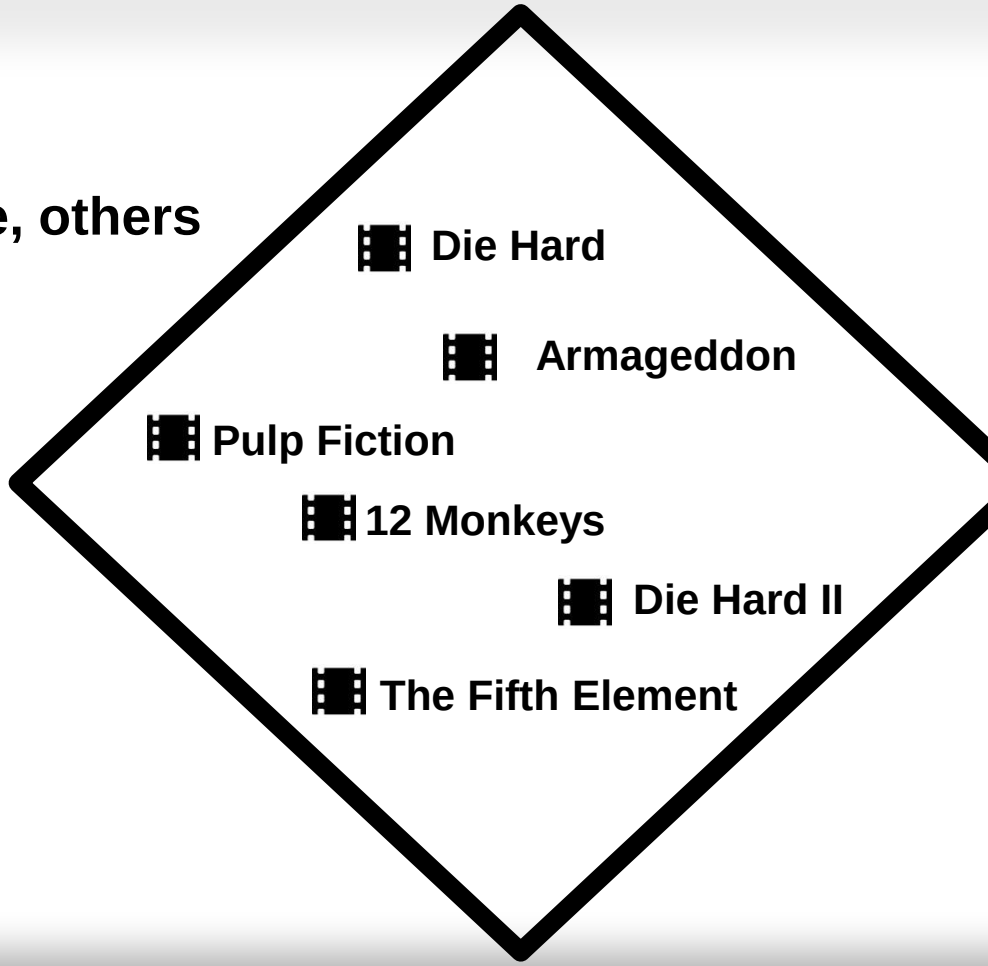
Video representation on server

Video representation on server



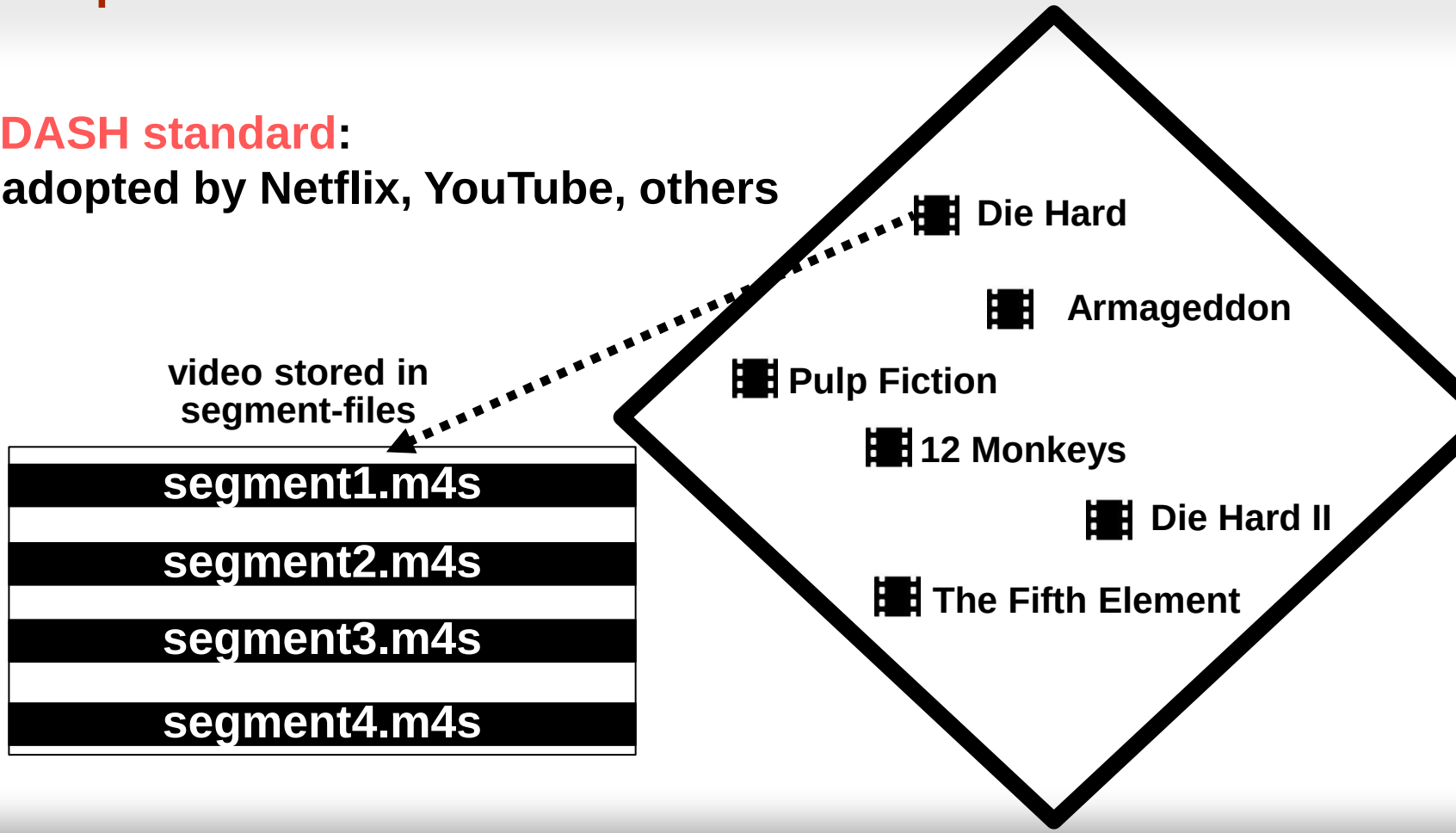
Video representation on server

MPEG-DASH standard:
widely adopted by Netflix, YouTube, others



Video representation on server

MPEG-DASH standard:
widely adopted by Netflix, YouTube, others



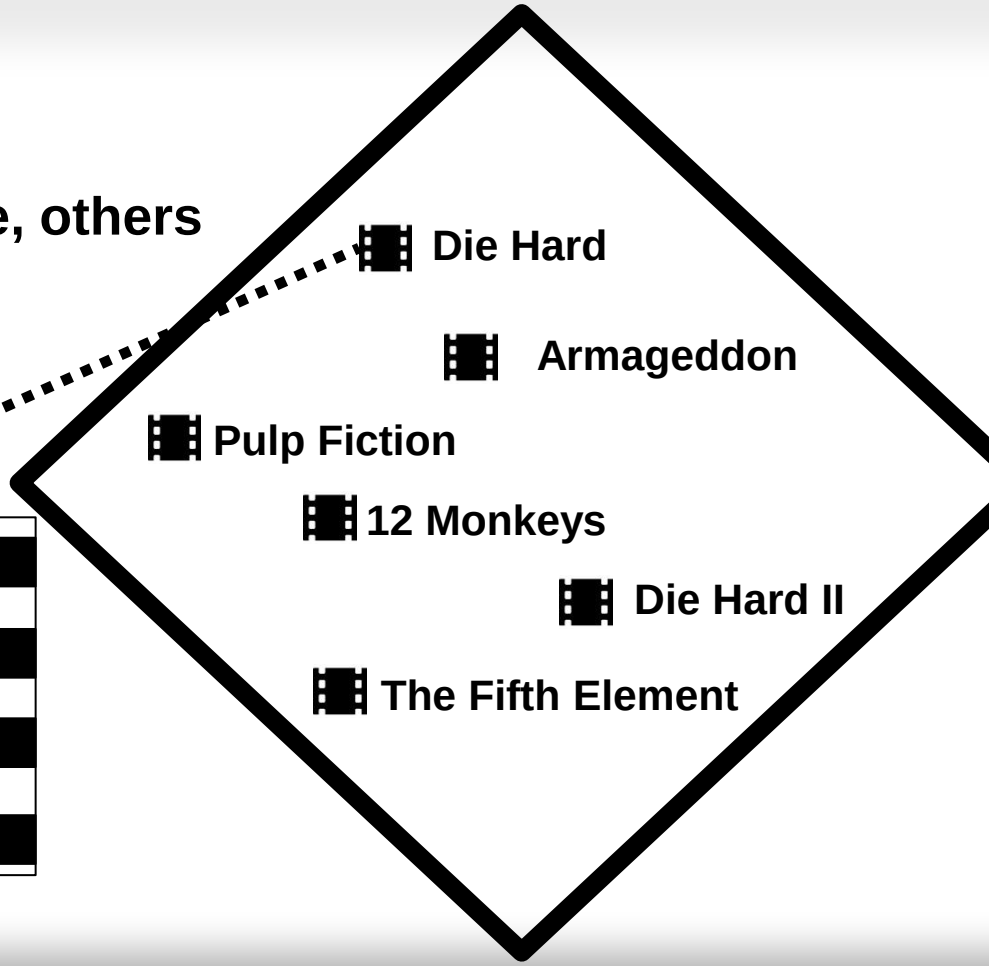
Video representation on server

MPEG-DASH standard:
widely adopted by Netflix, YouTube, others

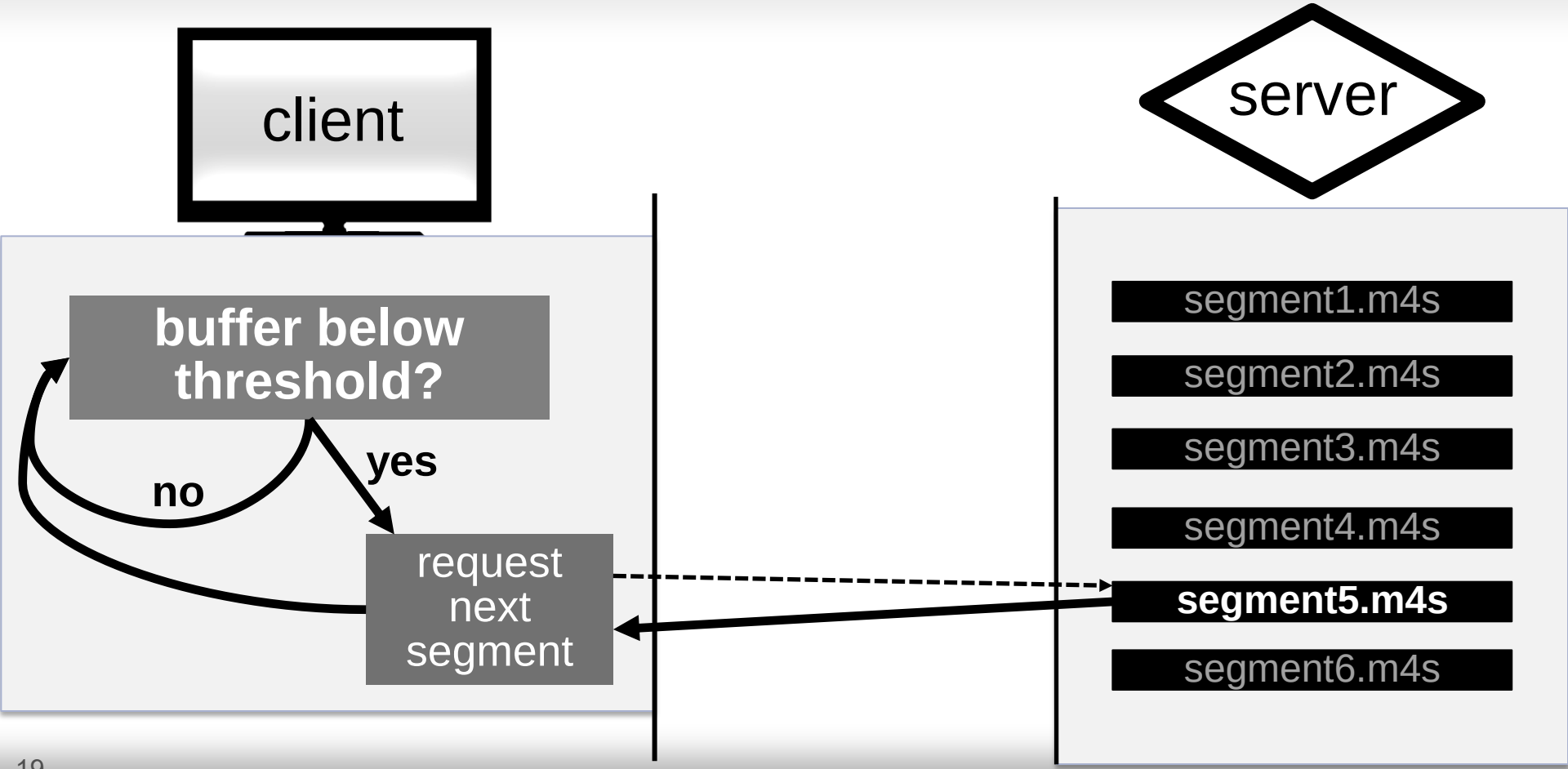
segment =
a few seconds of playback

video stored in segment-files

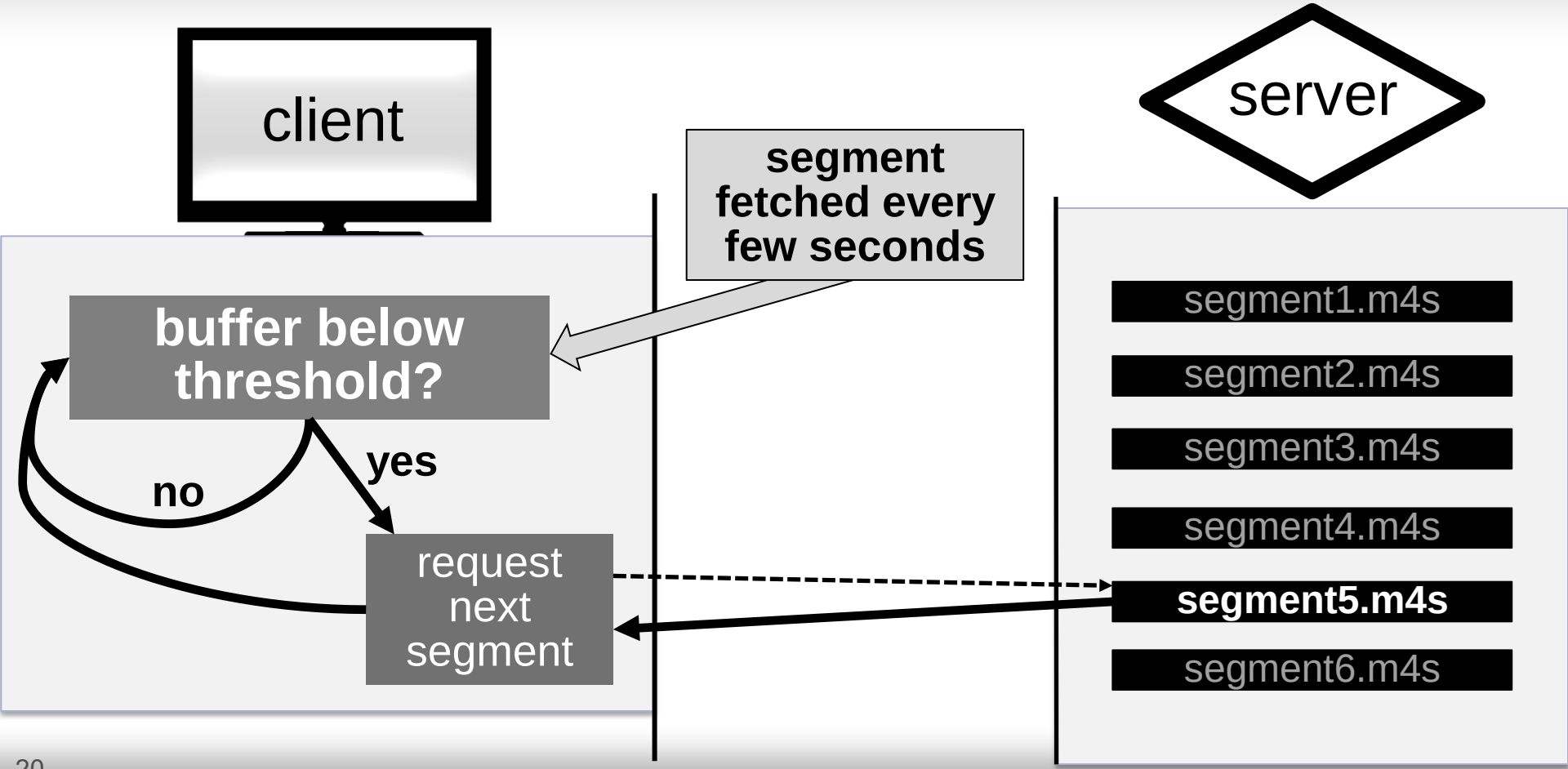
0-5sec	segment1.m4s
5-10sec	segment2.m4s
10-15sec	segment3.m4s
15-20sec	segment4.m4s



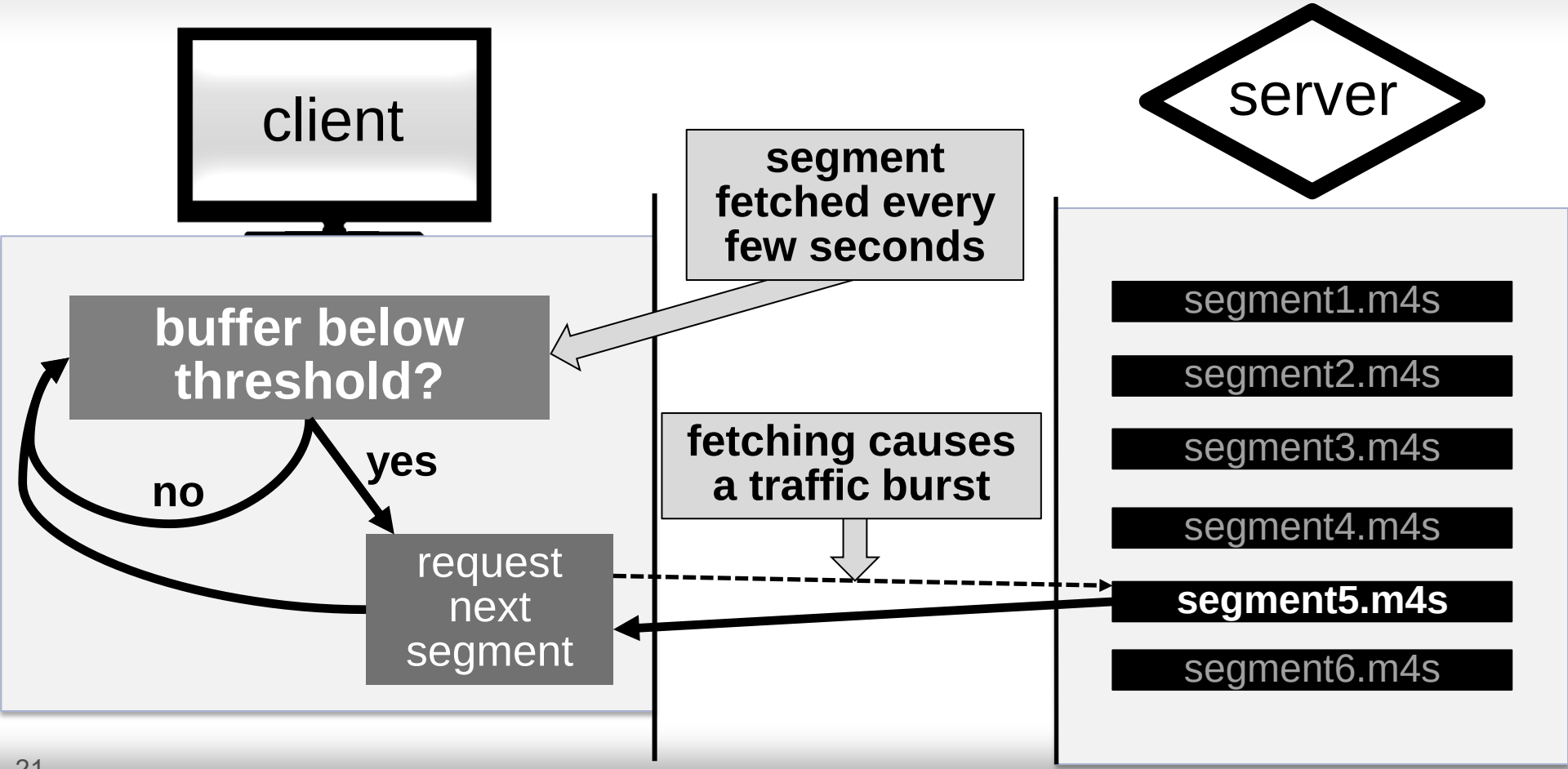
DASH client-server interaction (simplified)



DASH client-server interaction (simplified)

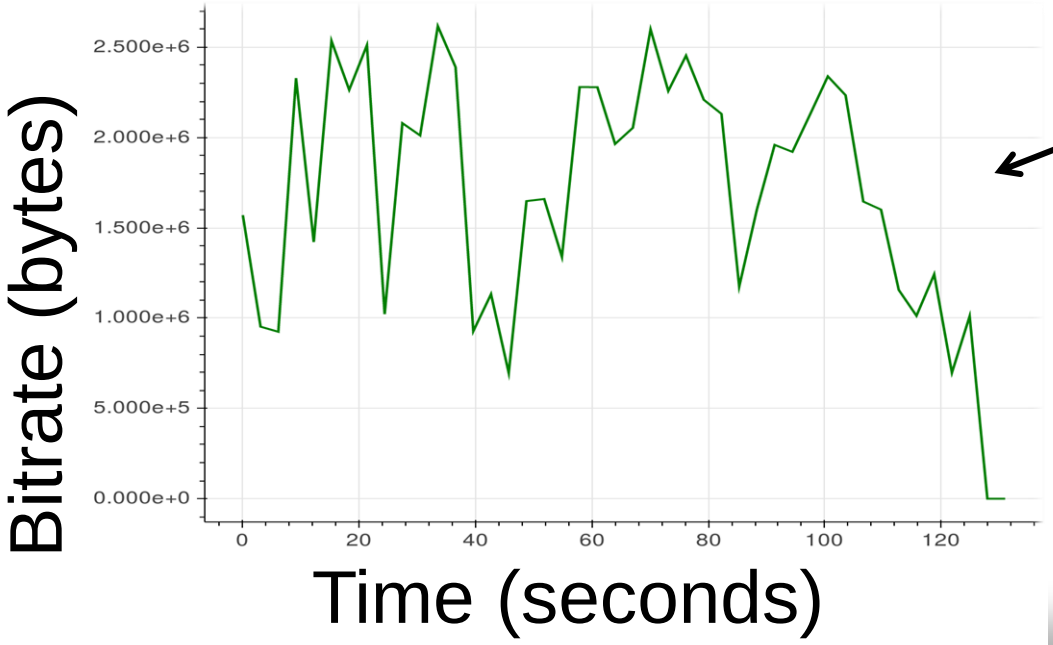


DASH client-server interaction (simplified)



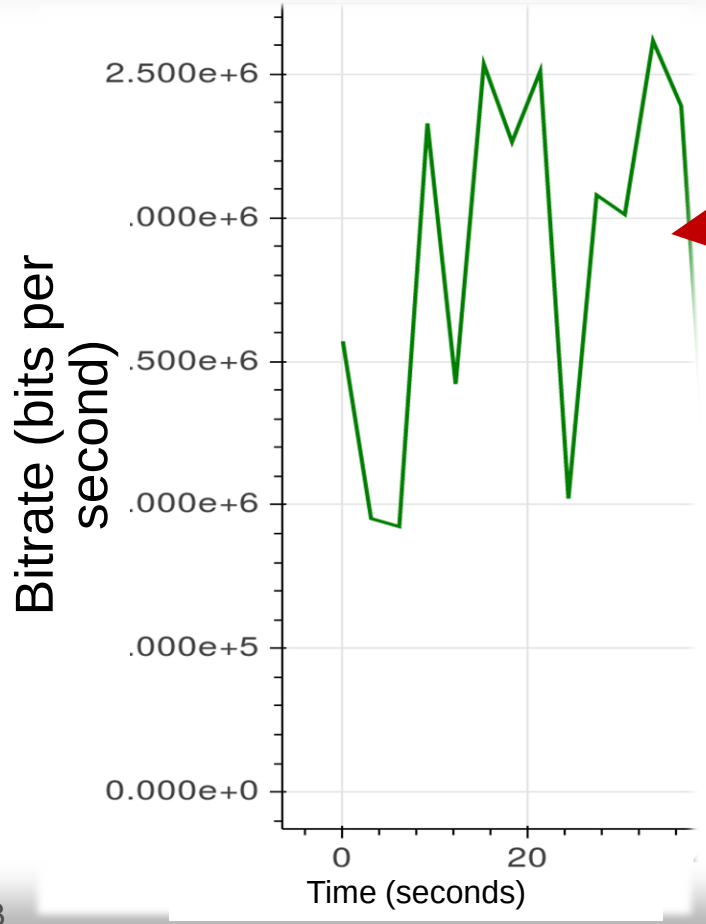
Variable bit rate encoding

Different video seconds require different amount of bytes to encode



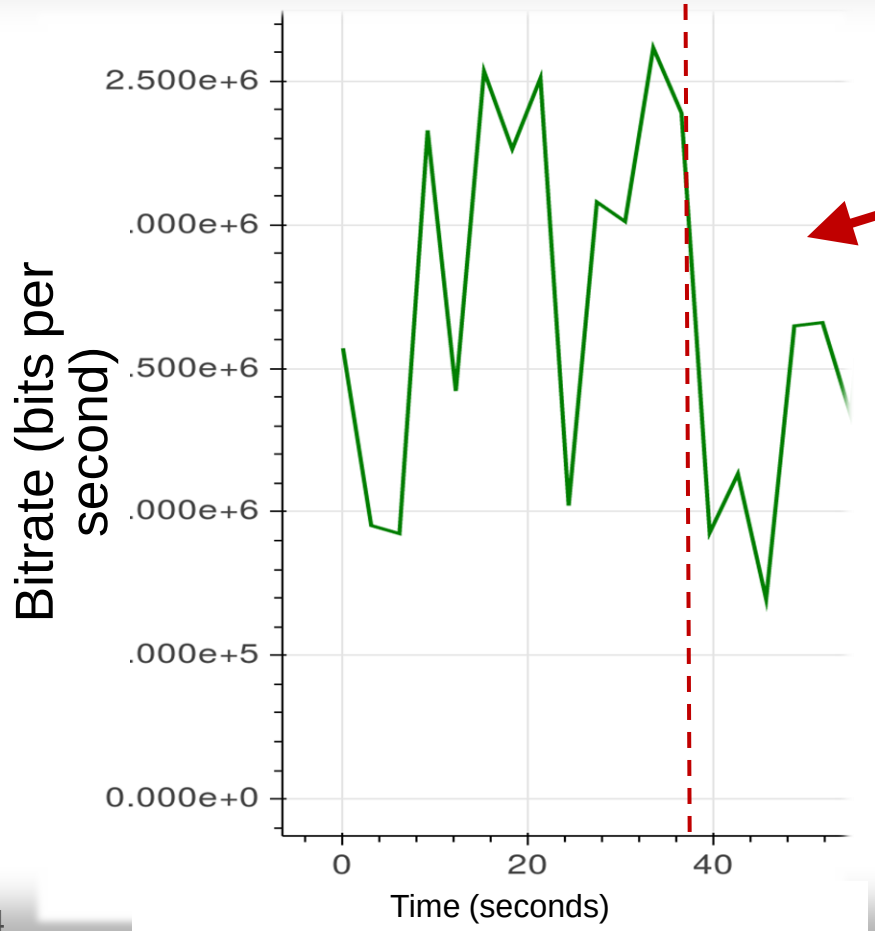
Iguana vs. Snakes VBR

Phases of Iguana vs Snakes in Bitrate



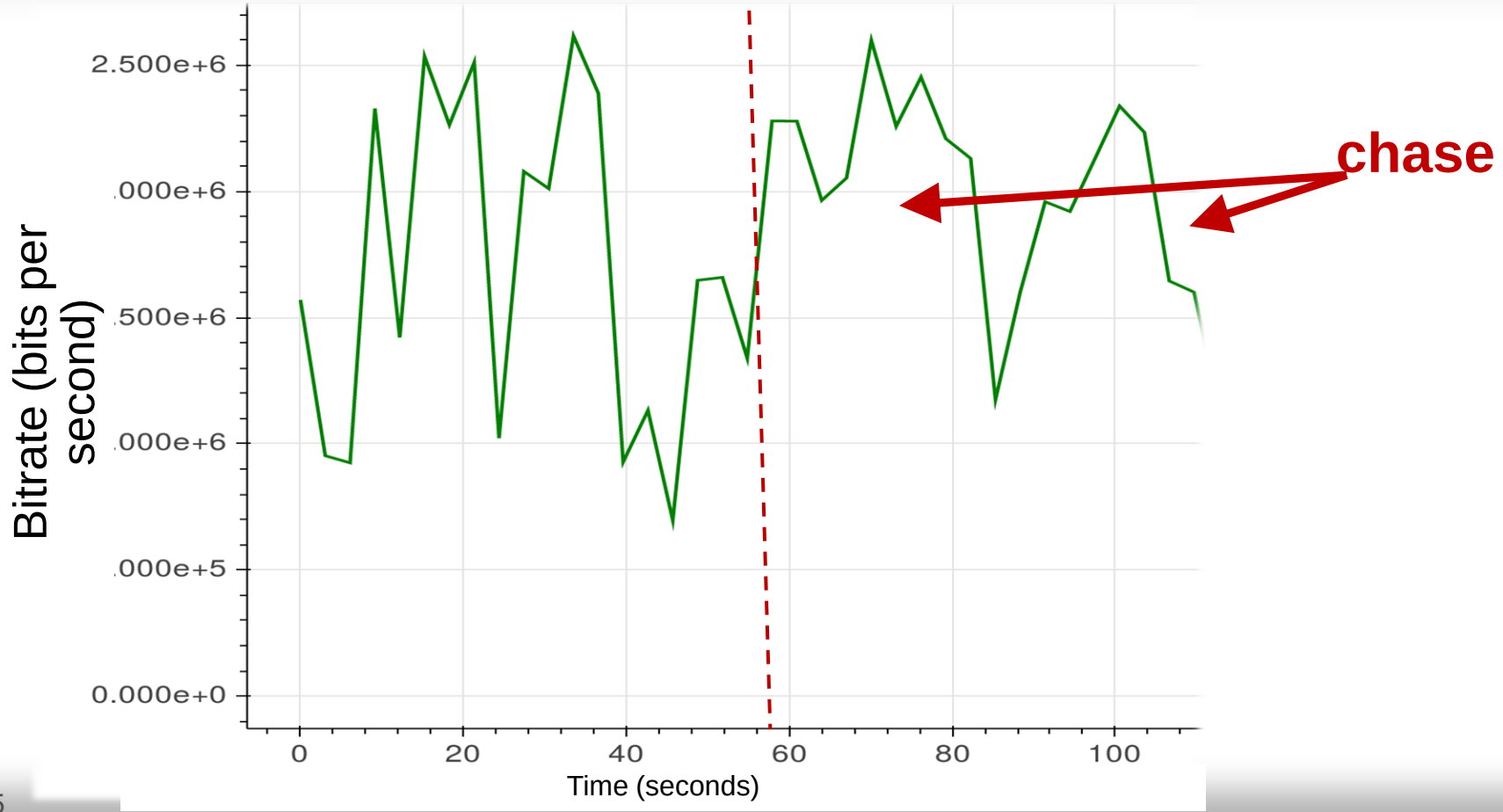
scenery, movement, tension rising

Phases of Iguana vs Snakes in Bitrate

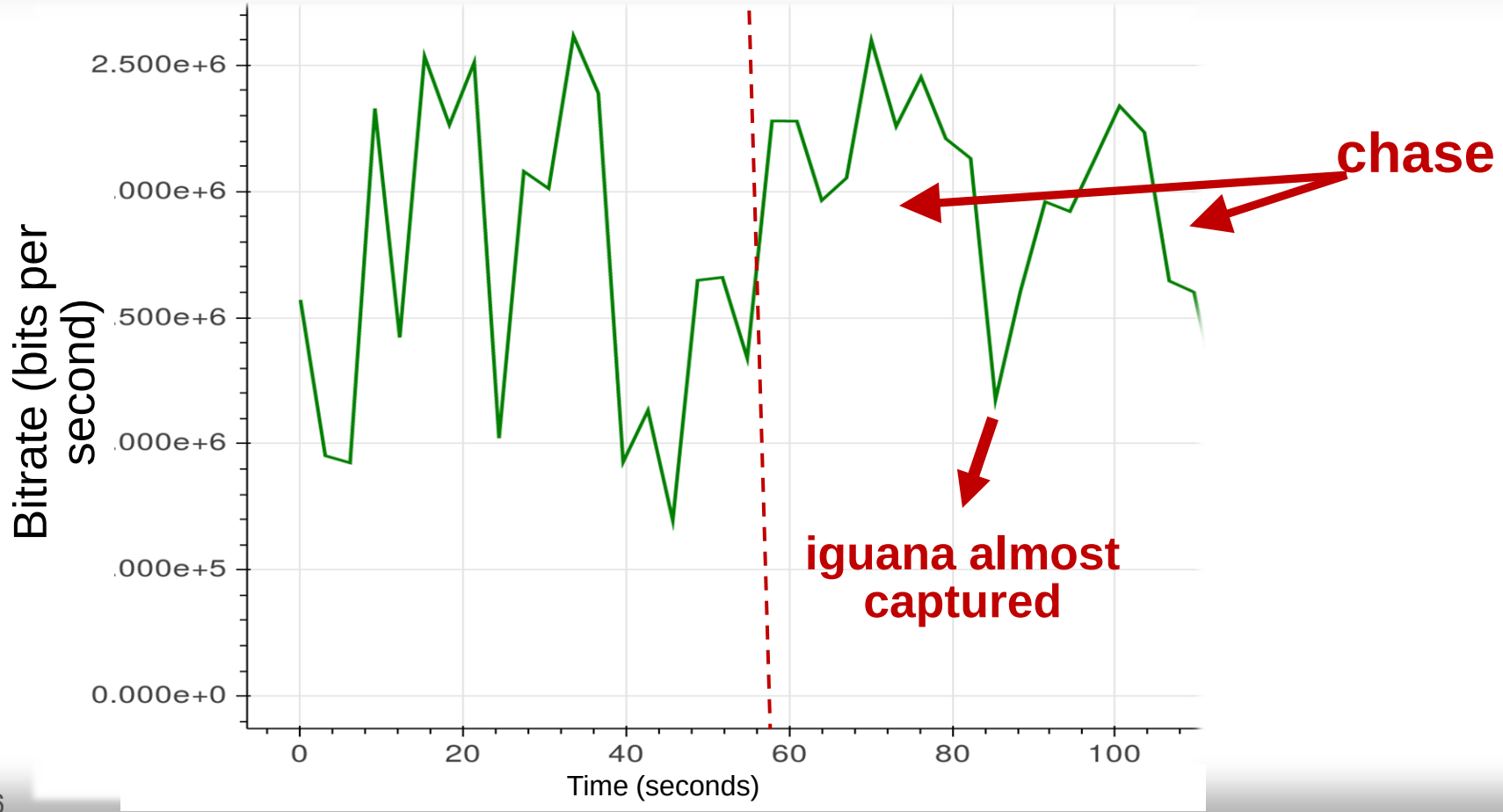


tension peaking, iguana is still

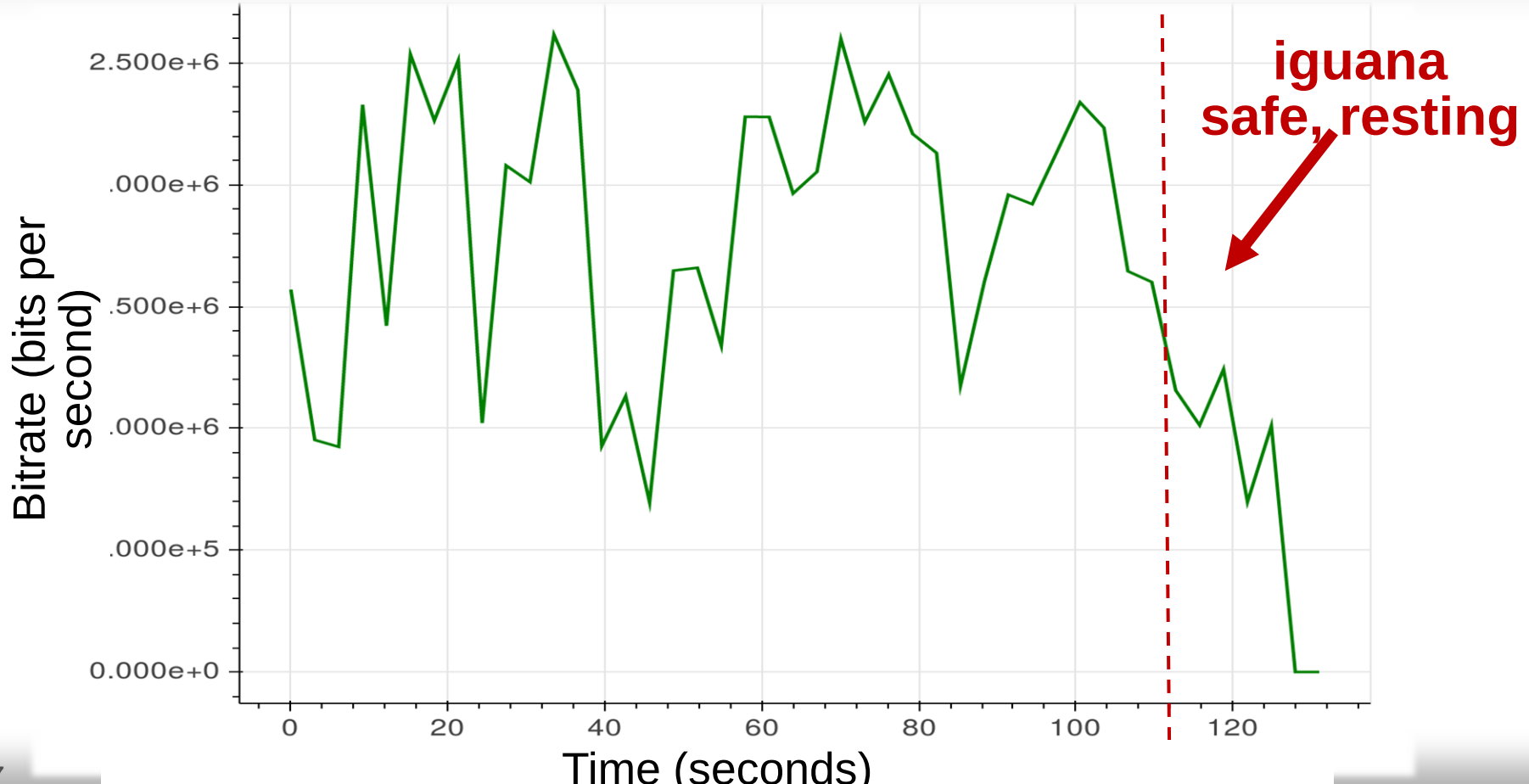
Phases of Iguana vs Snakes in Bitrate



Phases of Iguana vs Snakes in Bitrate

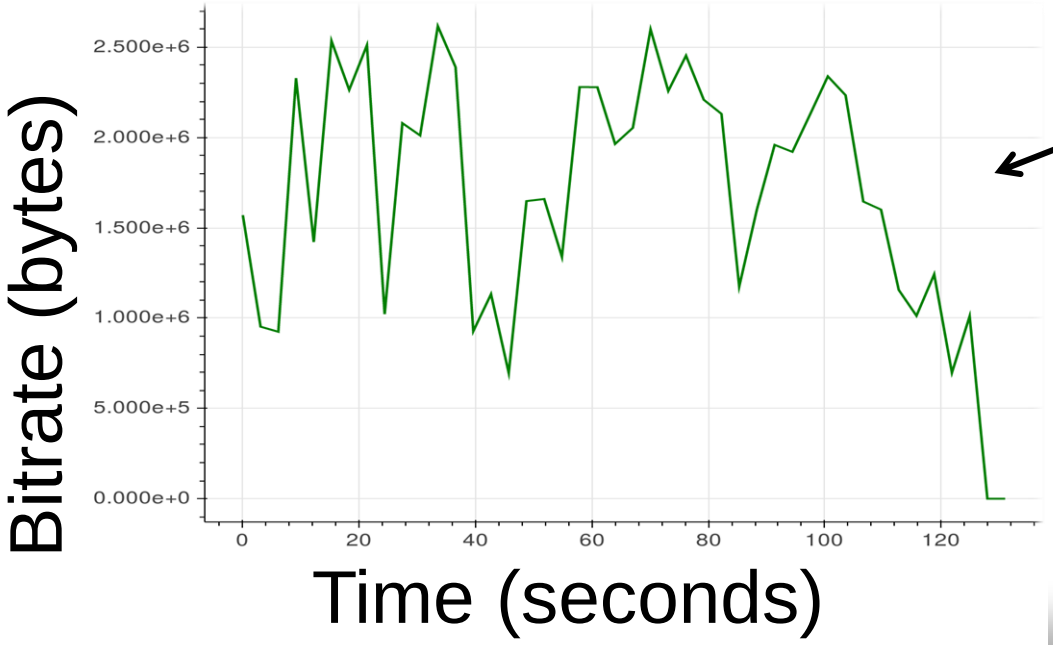


Phases of Iguana vs Snakes in Bitrate



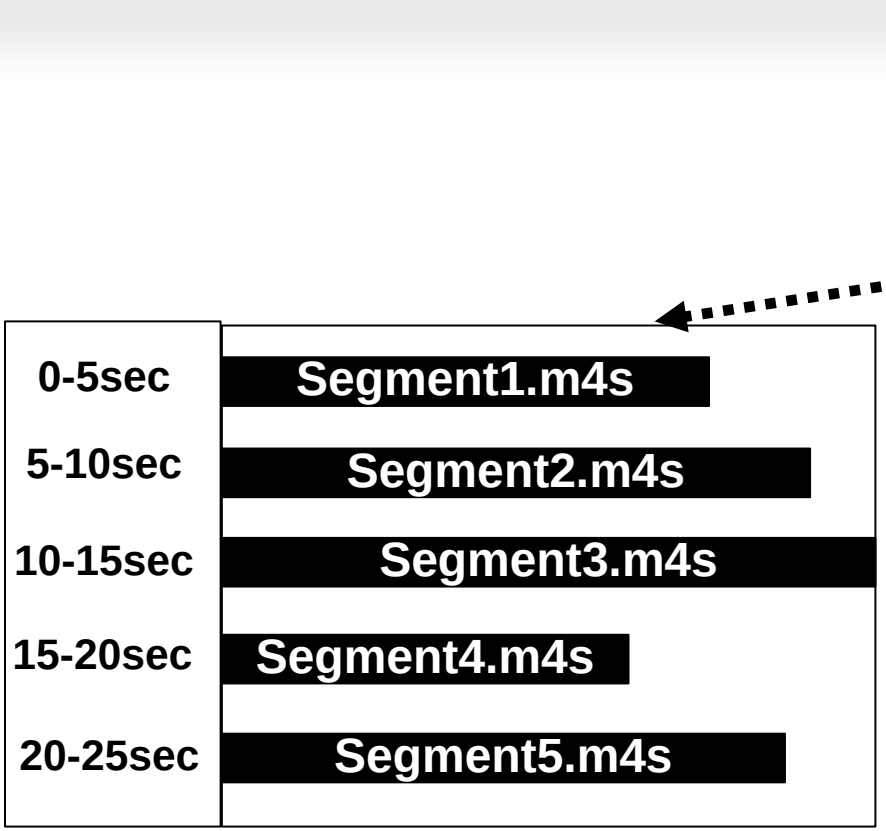
Variable bit rate encoding

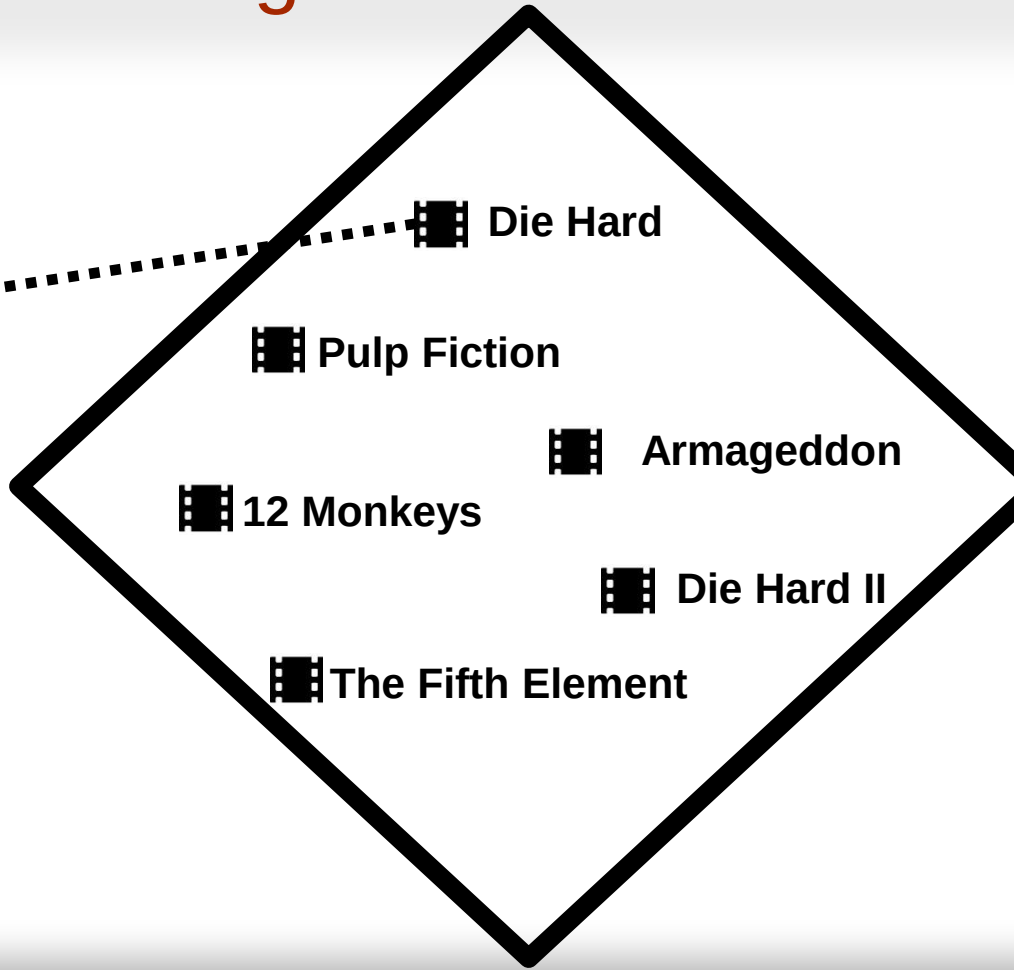
Different video seconds require different amount of bytes to encode



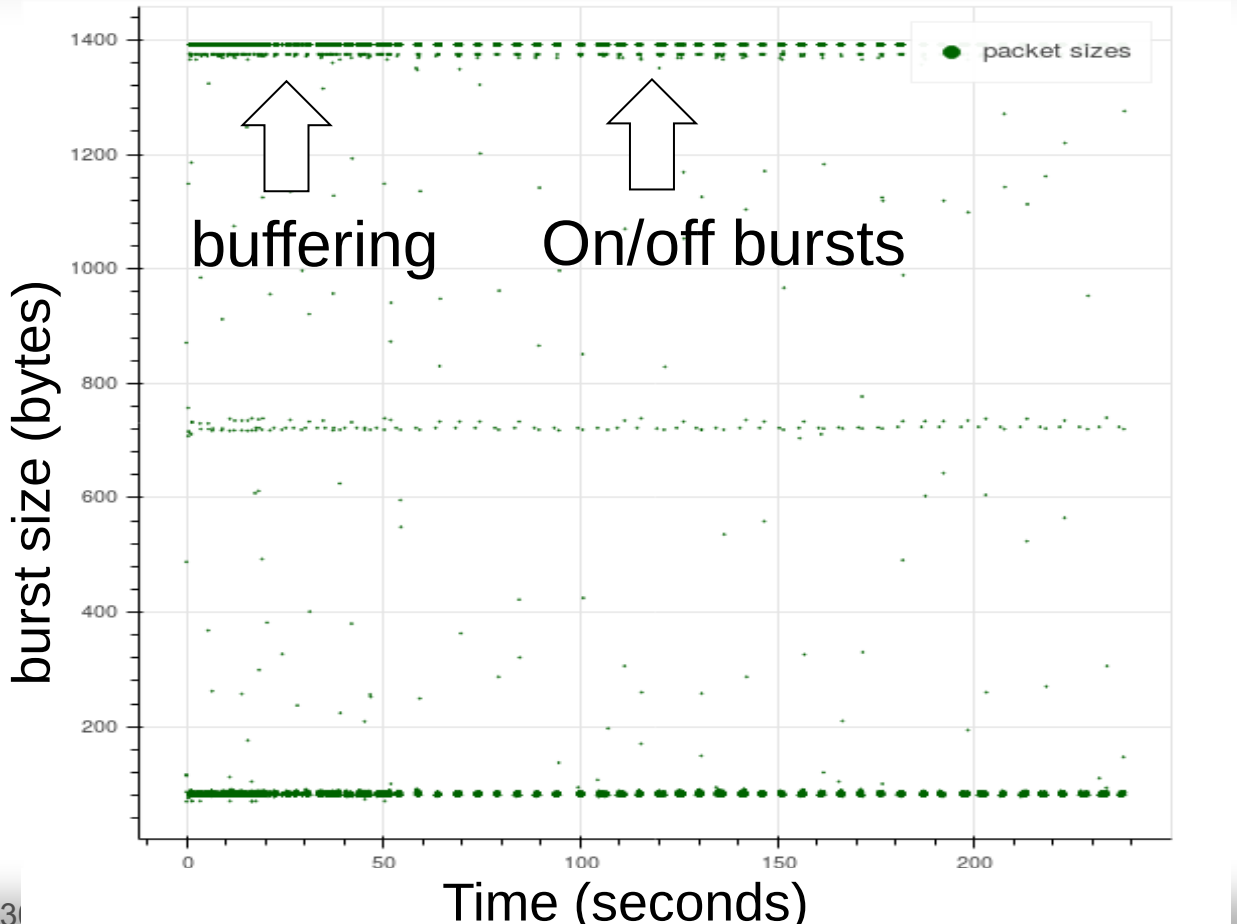
Iguana vs. Snakes VBR

Variable bit rate → variable segment size

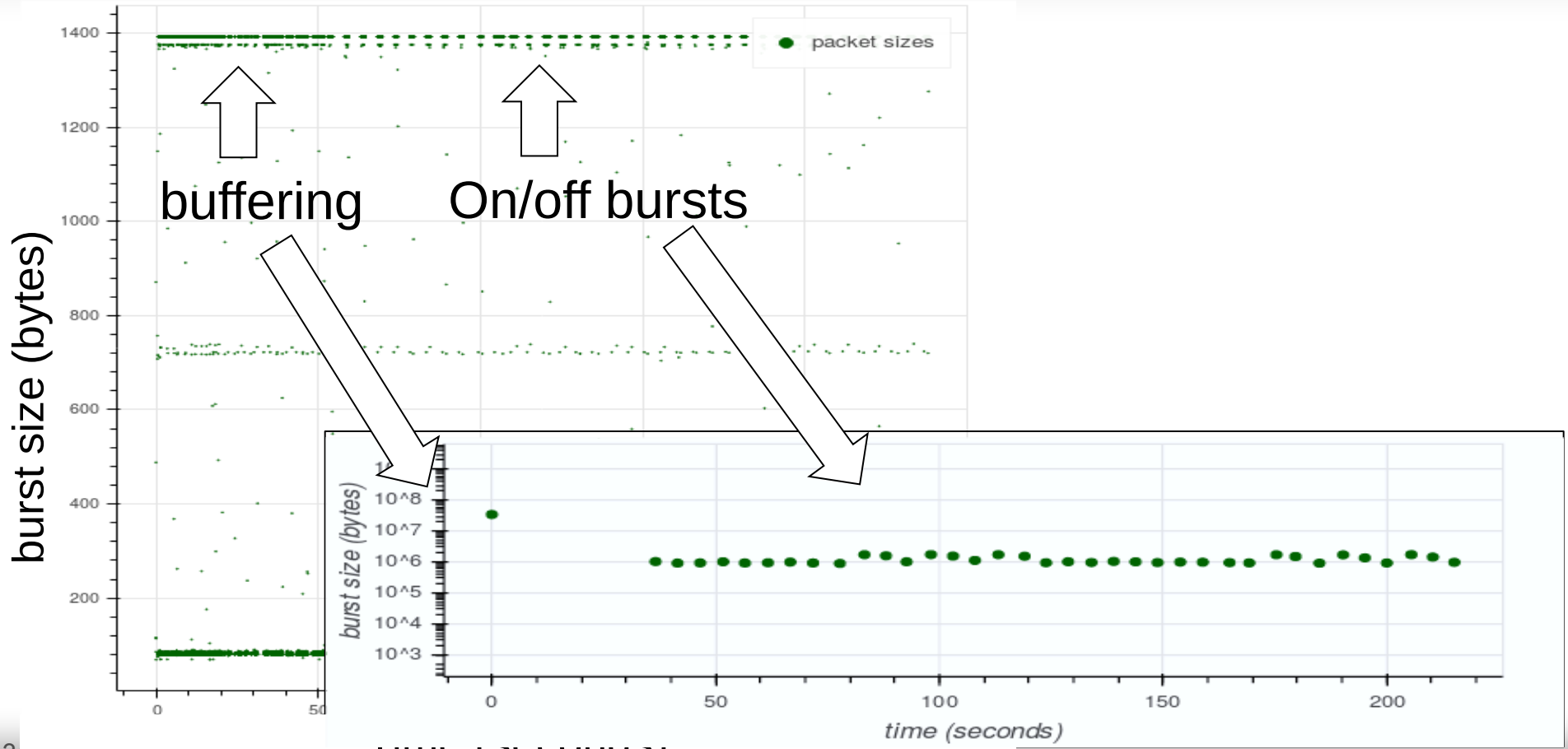




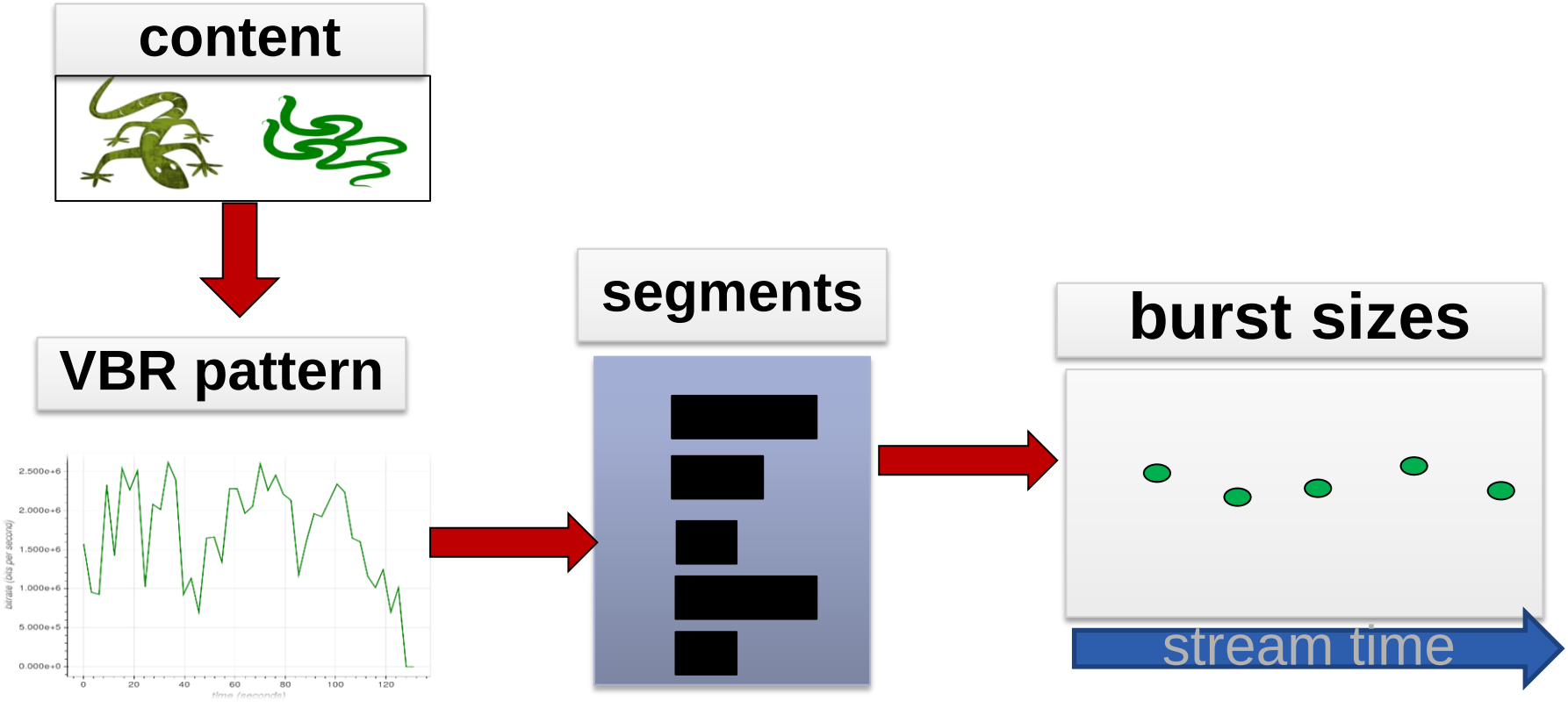
Variable segment size → variable burst size



Variable segment size → variable burst size



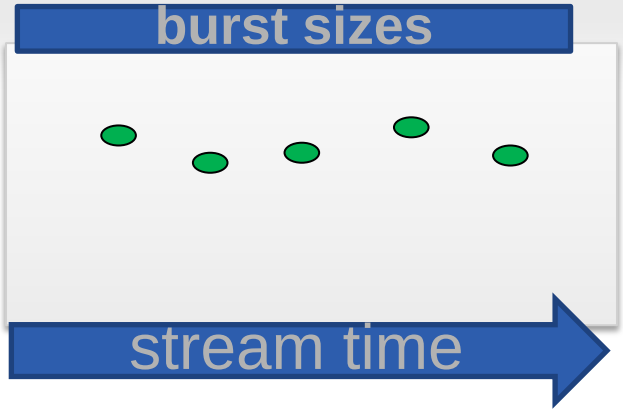
MPEG-DASH leak



From a leak to a fingerprint

Does the pattern of burst (segment) sizes uniquely characterize a title?

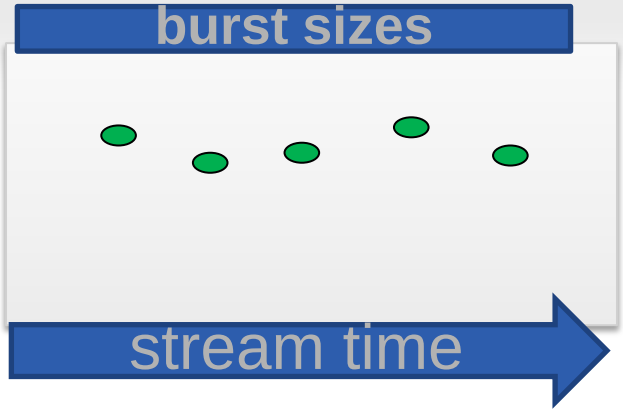
Can we learn a title's identifying pattern?



From a leak to a fingerprint

Does the pattern of burst (segment) sizes uniquely characterize a title?

Can we learn a title's identifying pattern?



Diversity: empirically measure pairwise distances for 3500 downloaded and segmented YouTube titles

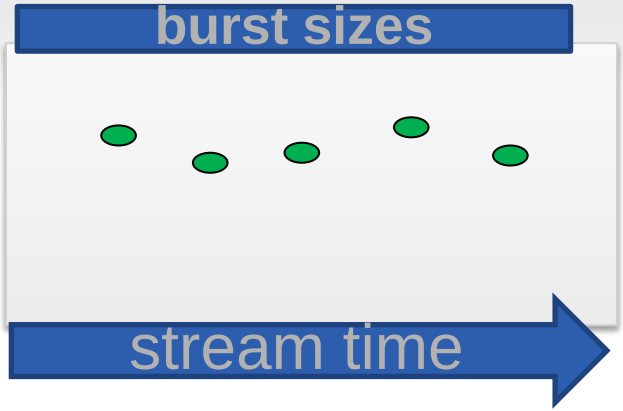
From a leak to a fingerprint

Does the pattern of burst (segment) sizes uniquely characterize a title?

Diversity: empirically measure pairwise distances for 3500 downloaded and segmented YouTube titles

Can we learn a title's identifying pattern?

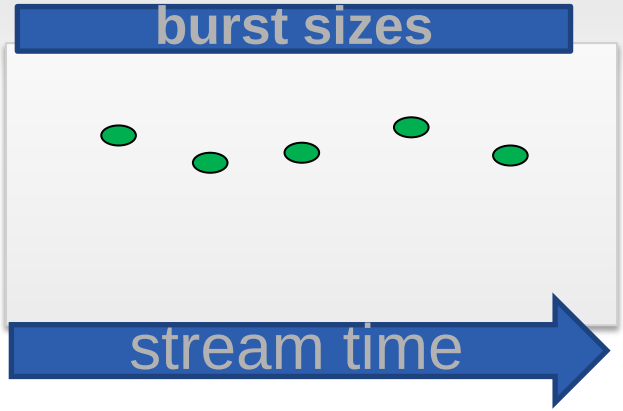
Consistency: empirically evaluate attacker's measurement error bound



From a leak to a fingerprint

Does the pattern of burst (segment) sizes uniquely characterize a title?

Can we learn a title's identifying pattern?



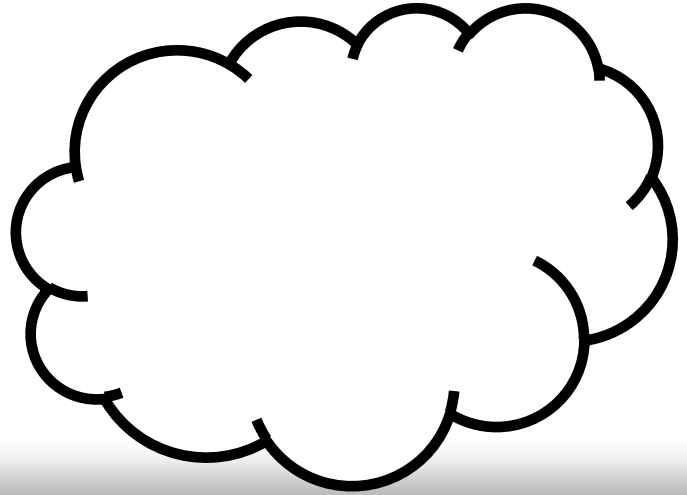
Diversity: empirically measure pairwise distances for 3500 downloaded and segmented YouTube titles

Consistency: empirically evaluate attacker's measurement error bound

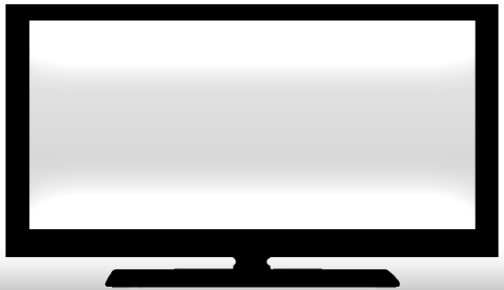
~20% of YouTube titles have fingerprints

Attack overview

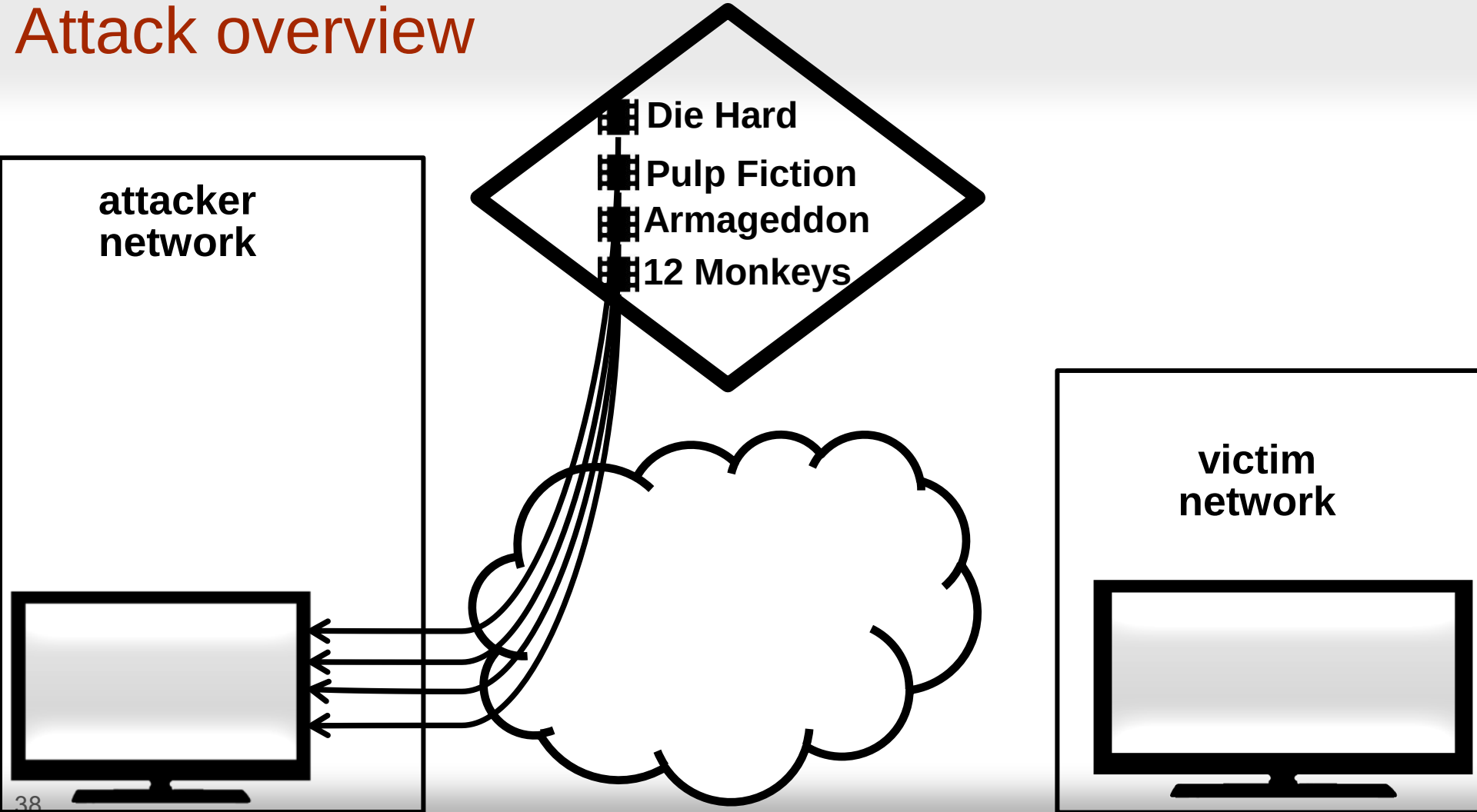
attacker network



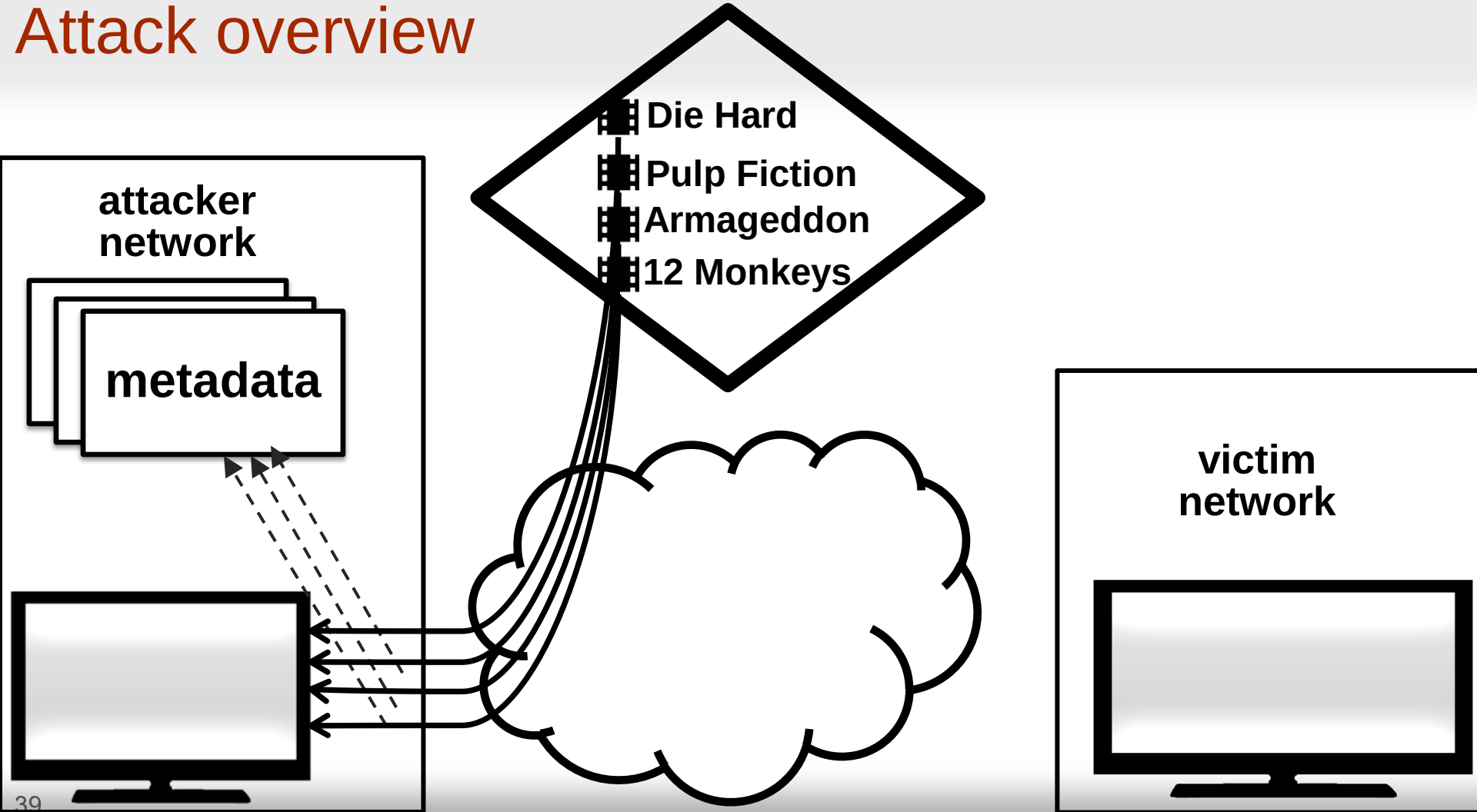
victim network



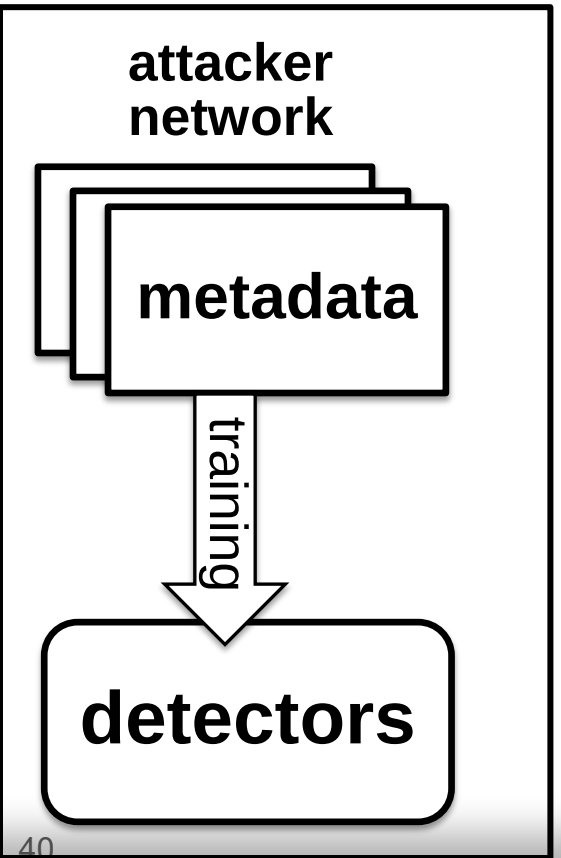
Attack overview

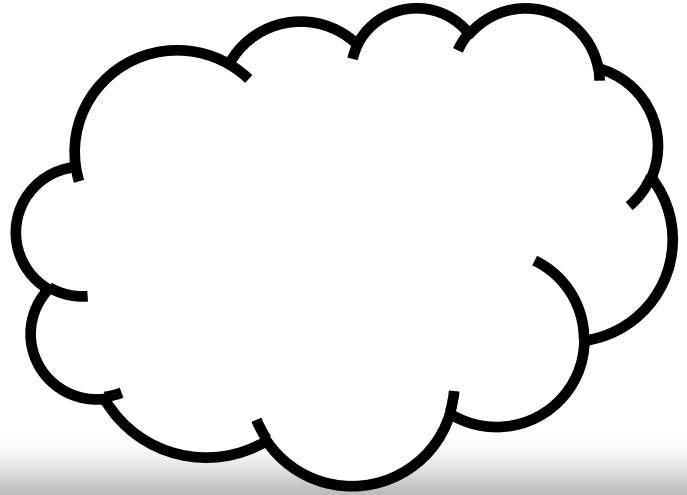


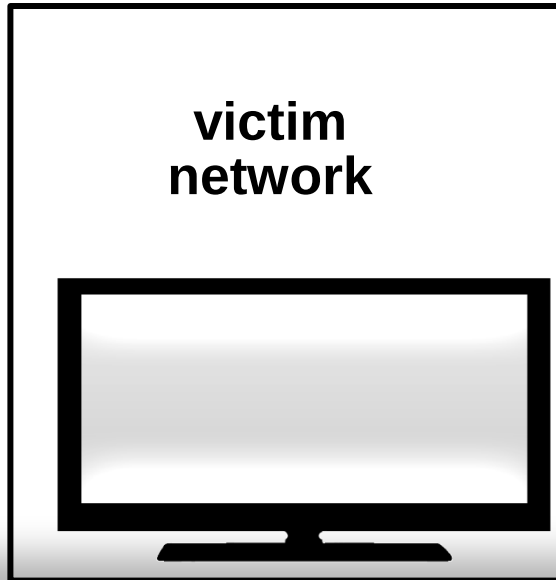
Attack overview



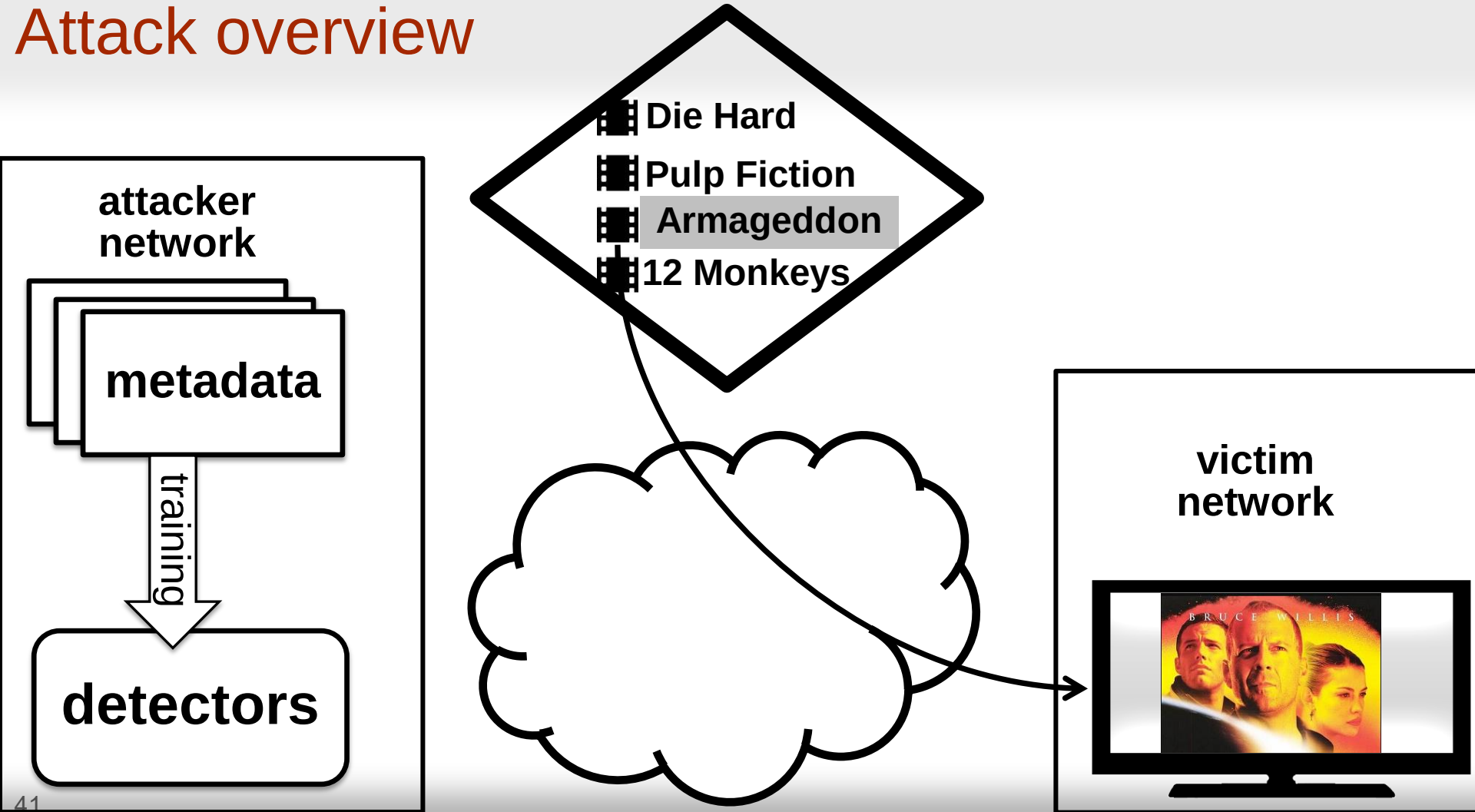
Attack overview



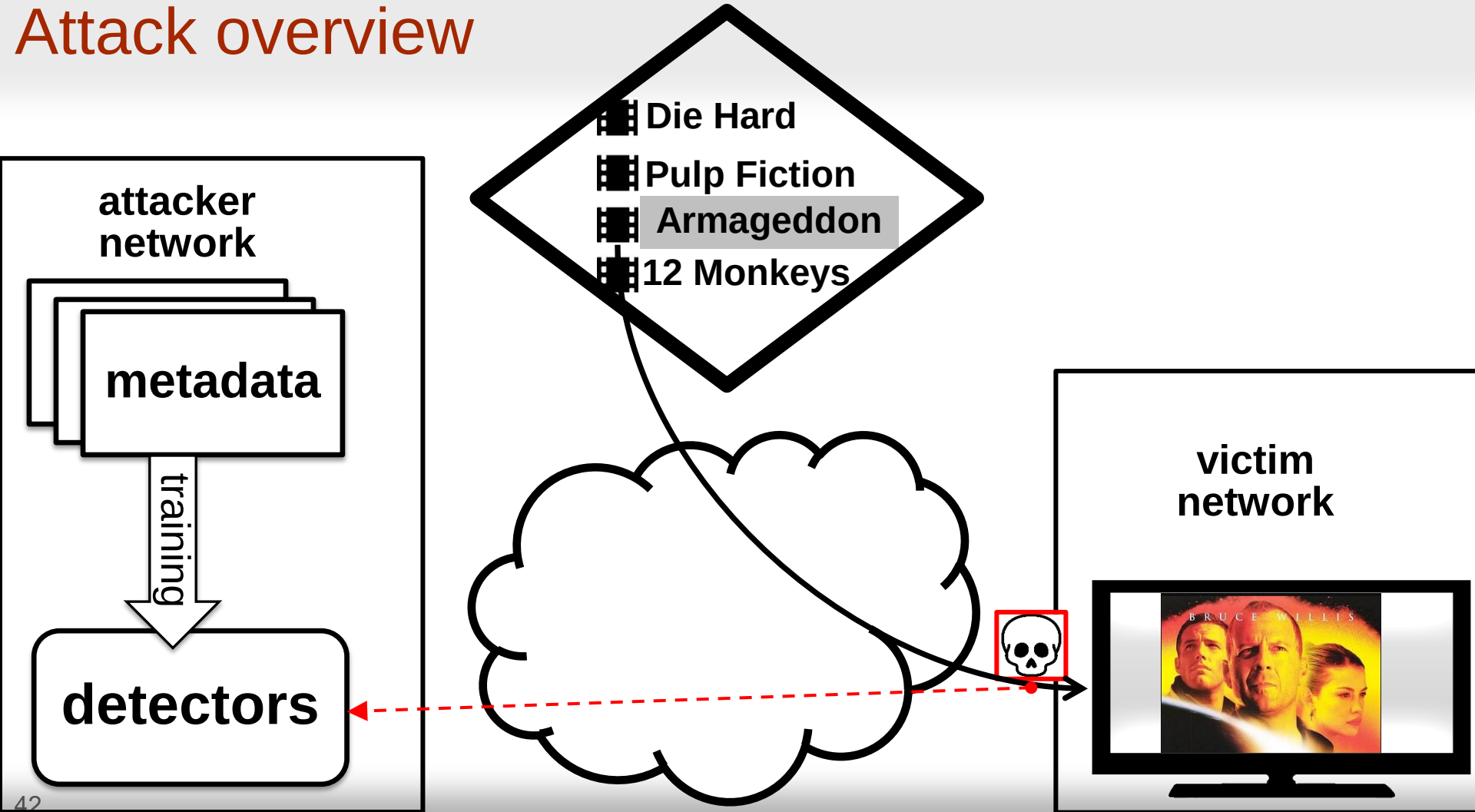




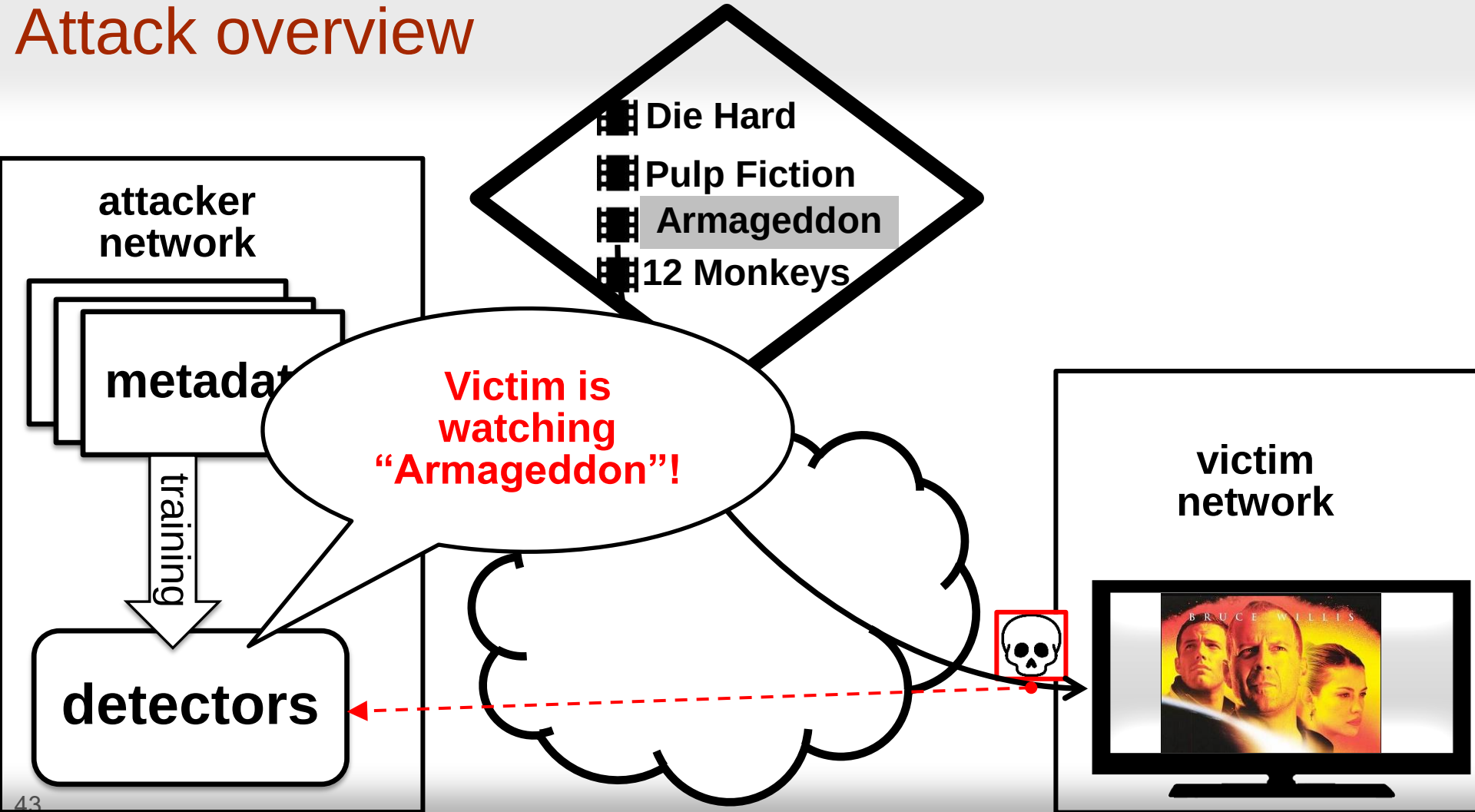
Attack overview



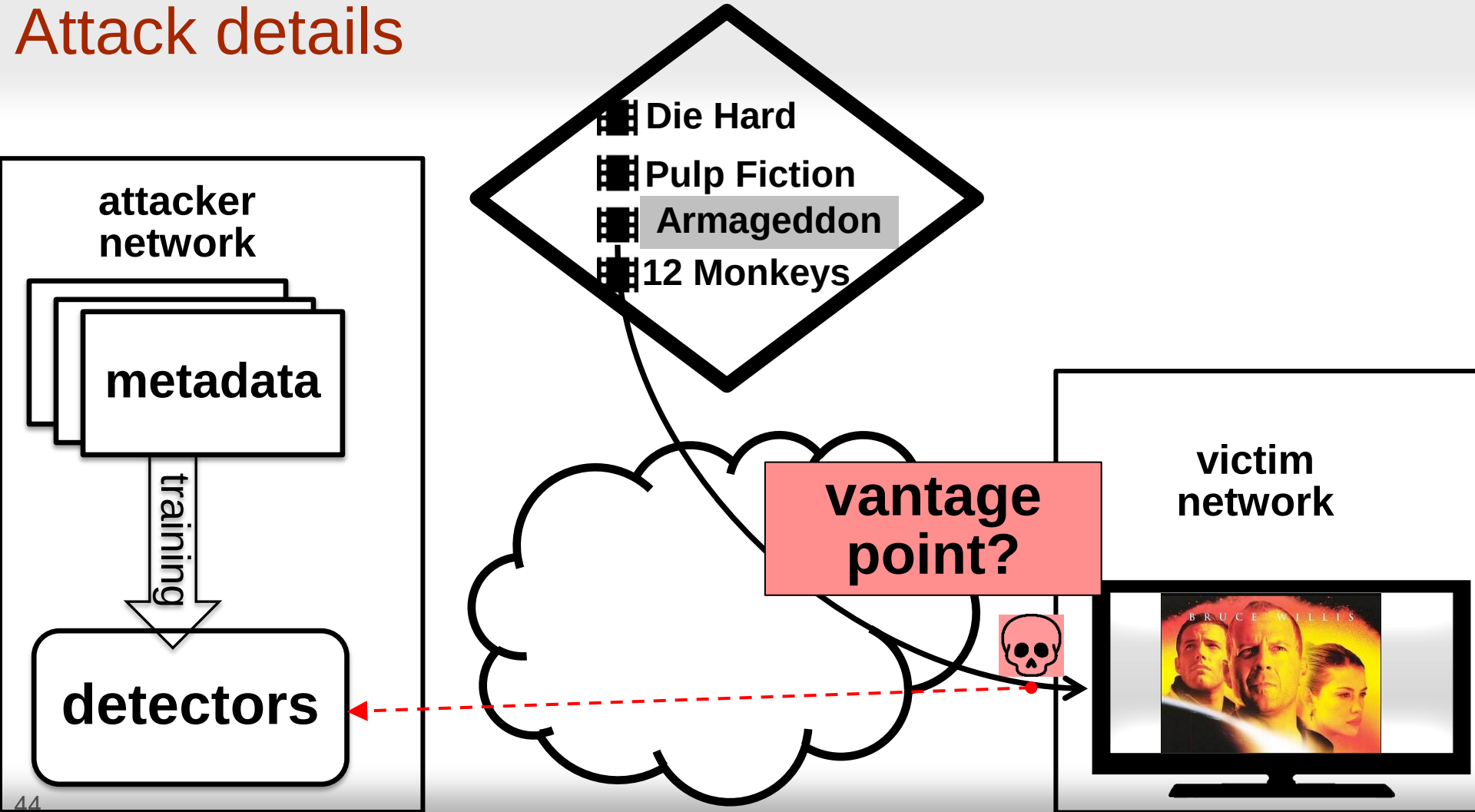
Attack overview



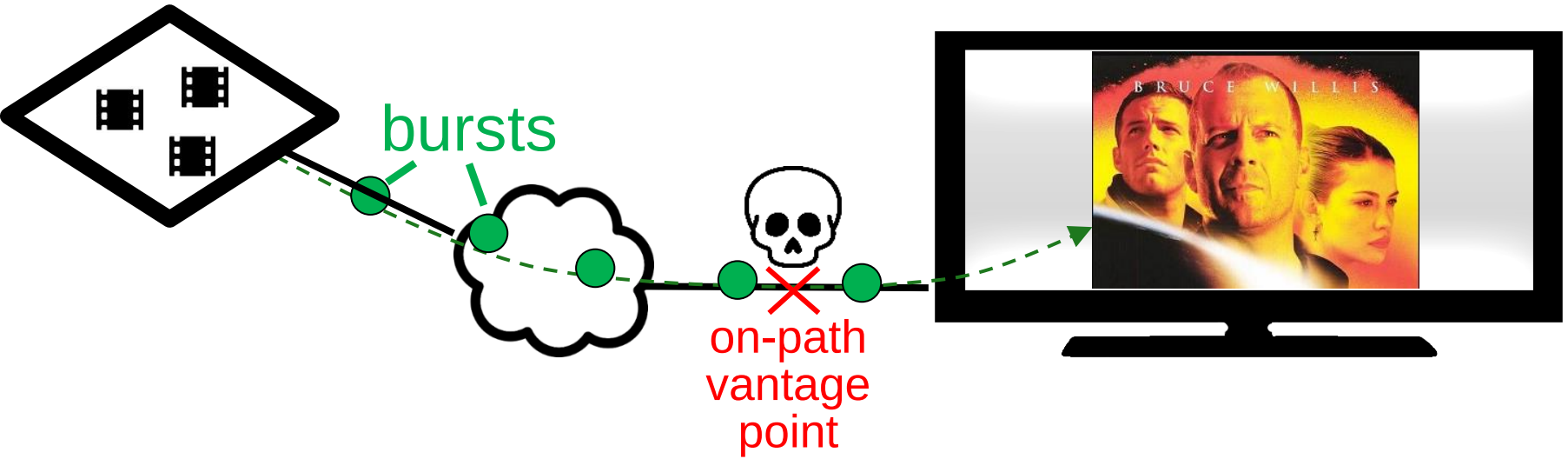
Attack overview



Attack details

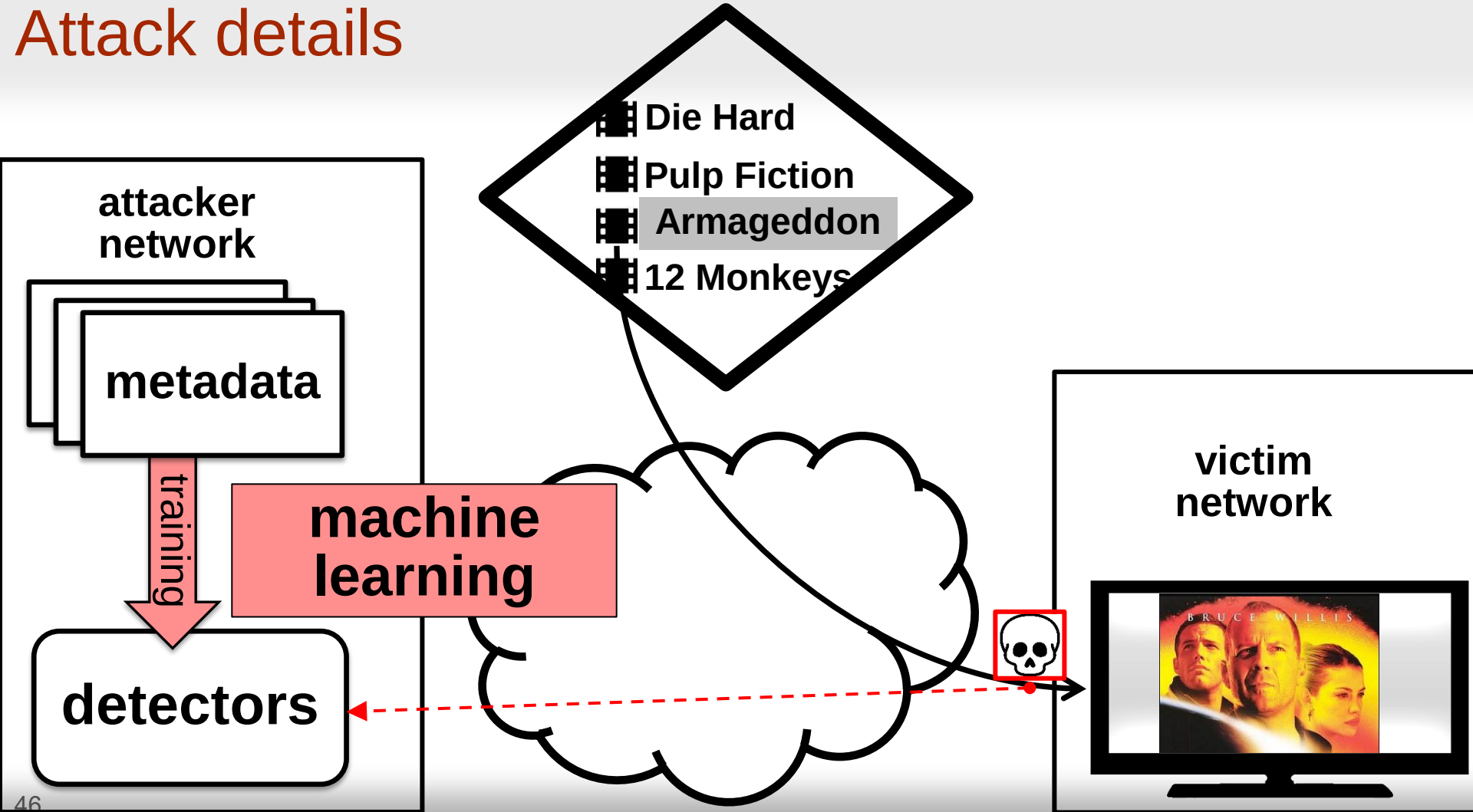


Scenario I: on-path attack



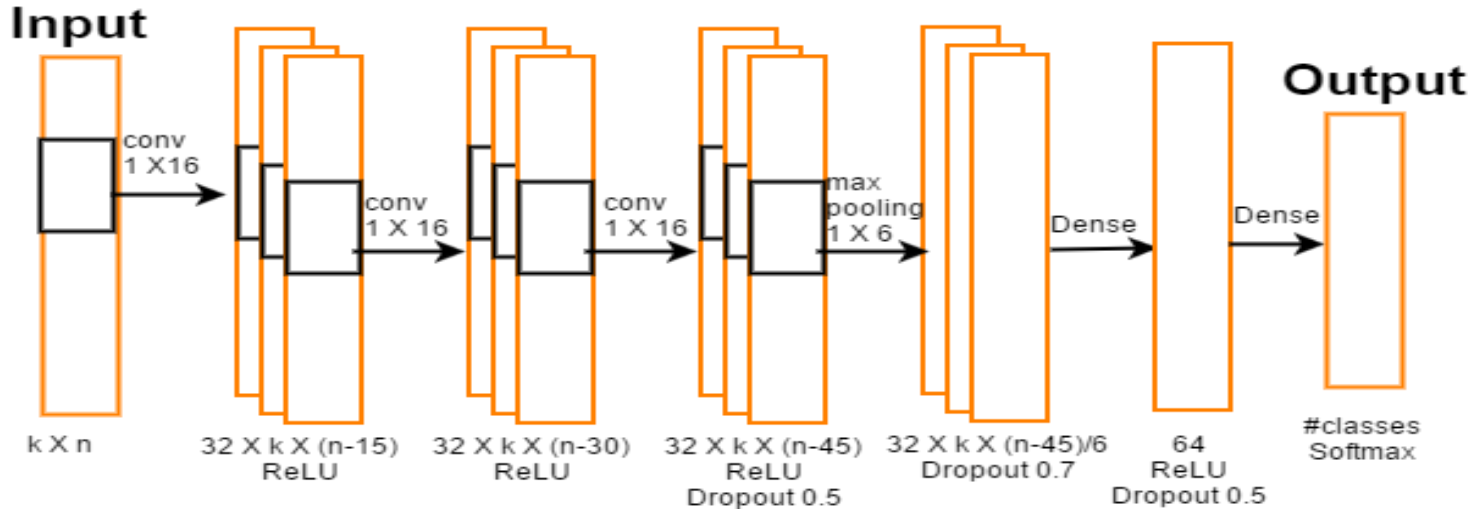
Wi-Fi access points, proxies, routers, enterprise or national network censors, ISPs

Attack details



Deep neural networks

- Very good at learning high-level concepts that are hard to express formally (e.g., “traffic traces are similar”)
- Existing NN architectures very accurate on classification and detection problems



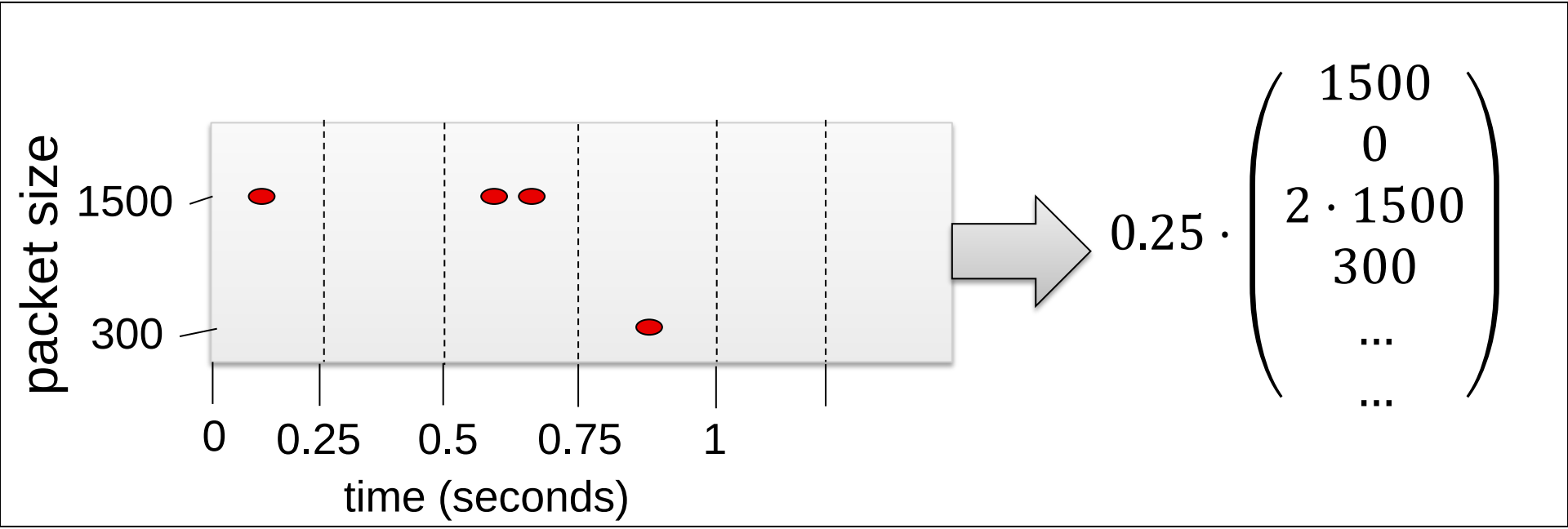
Advantages of neural networks

- Robust: can operate on noisy and coarse measurements
- Agnostic to protocol-specific attributes (e.g., QUIC vs. TLS)
- Can learn features other than burst patterns, e.g., arrival patterns of individual packets
- Can use multiple session representations, train on all at once

Features

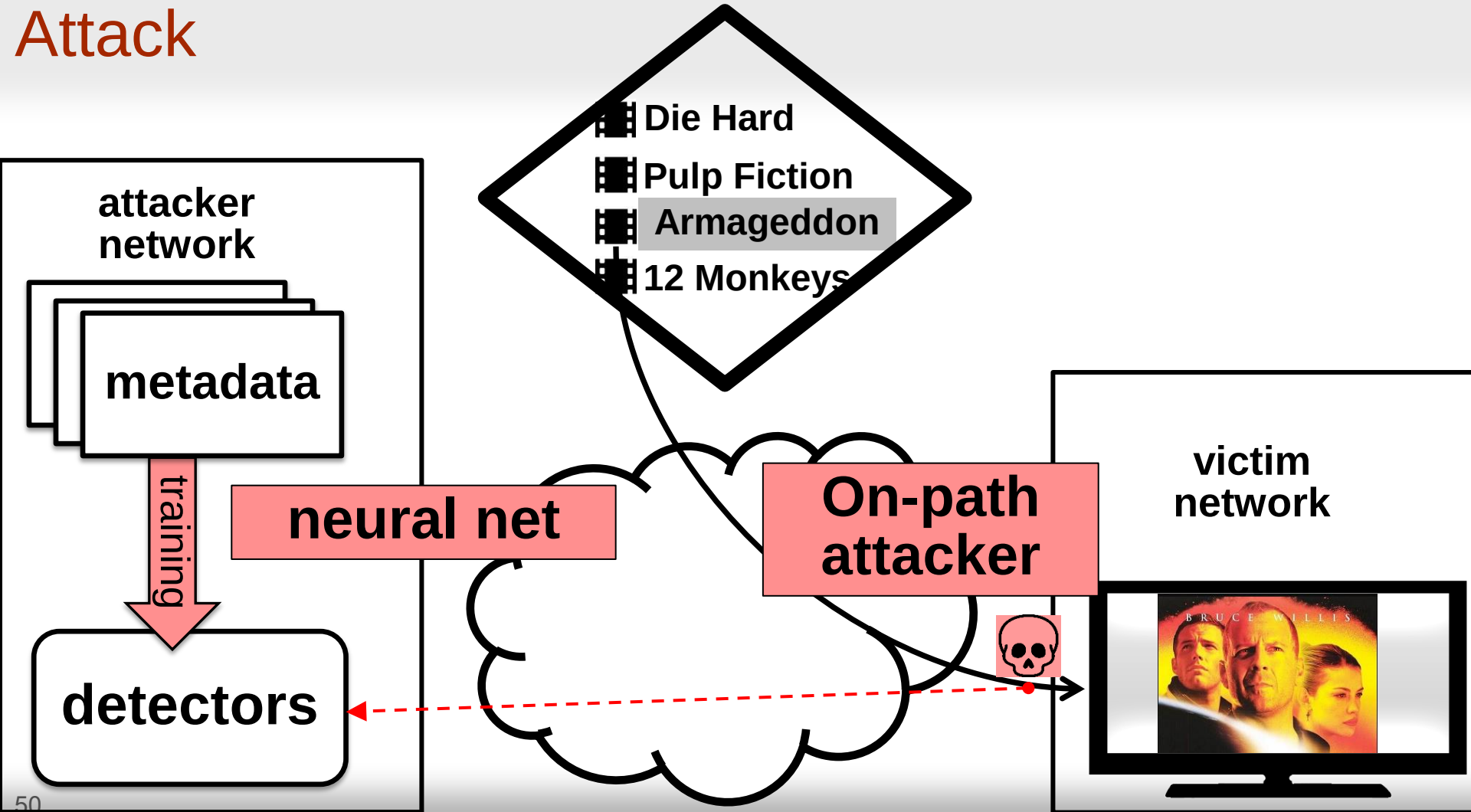
Each feature is a time-series, sampled at 0.25-second intervals

(example: bytes per second)



Features considered: downstream/upstream/total values of bytes per second, packet per second, average packet length, and burst sizes

Attack


Die Hard
Pulp Fiction
Armageddon
12 Monkeys
attacker network
metadata
training
detectors
neural net
On-path attacker
victim network
BRUCE WILLIS

Datasets and identification experiments

NETFLIX

100 titles
100 1-minute sessions

YouTube

18 titles
100 3-minute sessions
+
3500 sessions
of different other titles

amazon

10 titles
100 1.5-minute sessions

vimeo

10 titles
100 1-minute sessions

Datasets and identification experiments

NETFLIX

100 titles
100 1-minute sessions

100 classes

YouTube

18 titles
100 3-minute sessions
+
3500 sessions
of different other titles

amazon

10 titles
100 1.5-minute sessions

vimeo

10 titles
100 1-minute sessions

Datasets and identification experiments

NETFLIX

100 titles
100 1-minute sessions

100 classes

YouTube

open-world identification

18 titles
100 3-minute sessions
+
3500 sessions
of different other titles

18+1=19 classes

amazon

10 titles
100 1.5-minute sessions

vimeo

10 titles
100 1-minute sessions

Datasets and identification experiments

NETFLIX

100 titles
100 1-minute sessions

100 classes

YouTube

open-world identification

18 titles
100 3-minute sessions
+
3500 sessions
of different other titles

18+1=19 classes

amazon

10 titles
100 1.5-minute sessions

10 classes

vimeo

10 titles
100 1-minute sessions

10 classes

Datasets and identification experiments

NETFLIX

100 titles
100 1-minute sessions

100 classes
98.5% accuracy

YouTube

18 titles
100 3-minute sessions
+
3500 sessions
of different other titles

open-world identification
18**+1**=19 classes
99.5% accuracy

amazon

10 titles
100 1.5-minute sessions

10 classes
92.5% accuracy

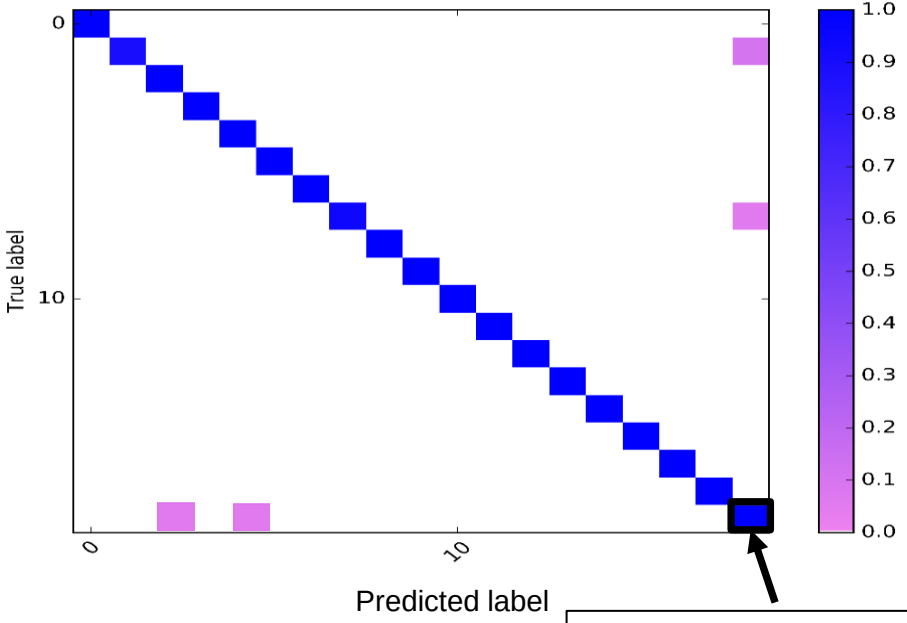
vimeo

10 titles
100 1-minute sessions

10 classes
98.6% accuracy

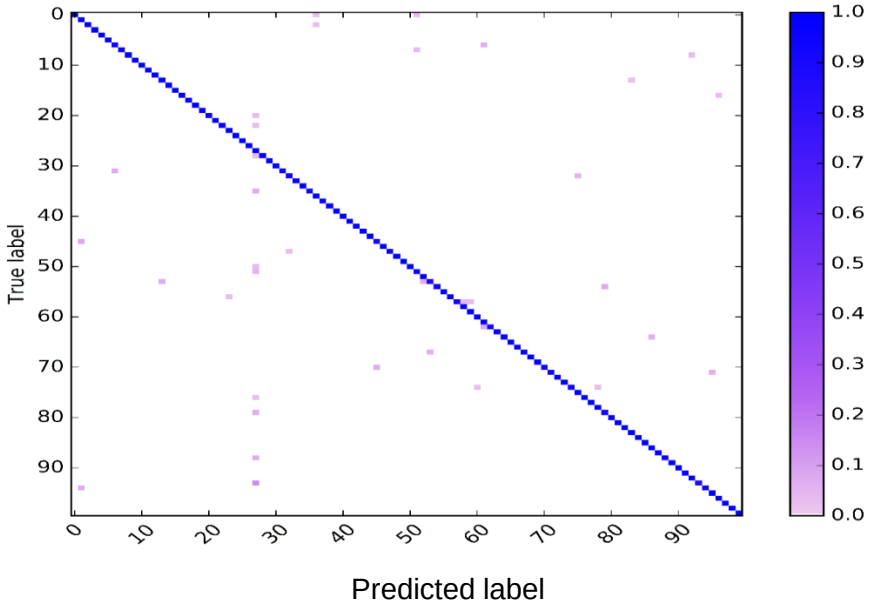
Empirical results: confusion matrices

YouTube
(feature: total burst size)



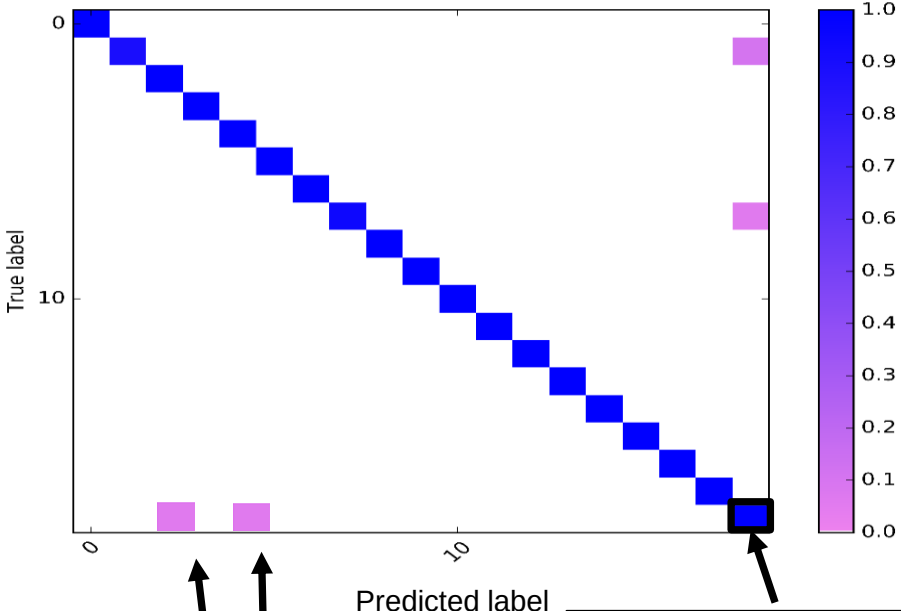
“unknown”
class, 3500 samples

Netflix
(feature: total burst size)



Empirical results: confusion matrices

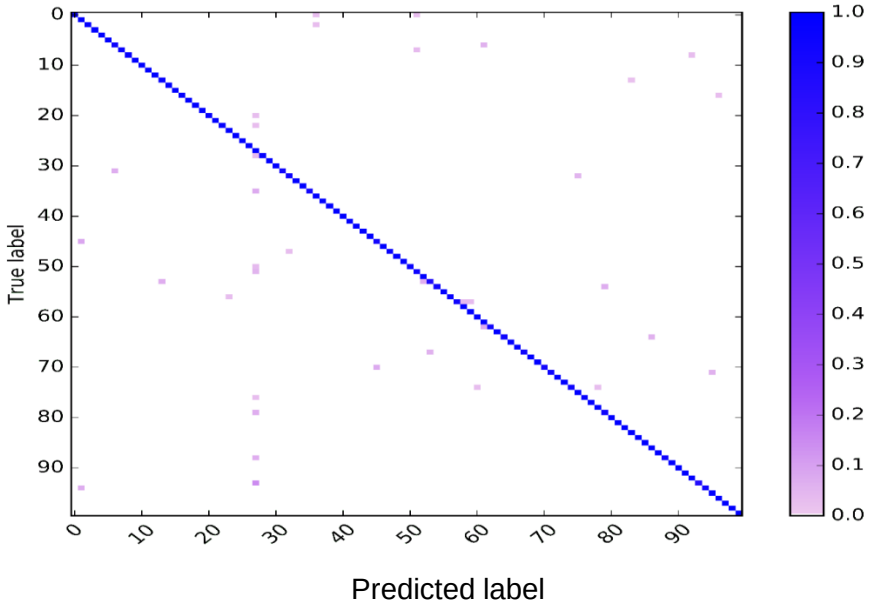
YouTube
(feature: total burst size)



Exactly 2 false positives

"unknown" class, 3500 samples

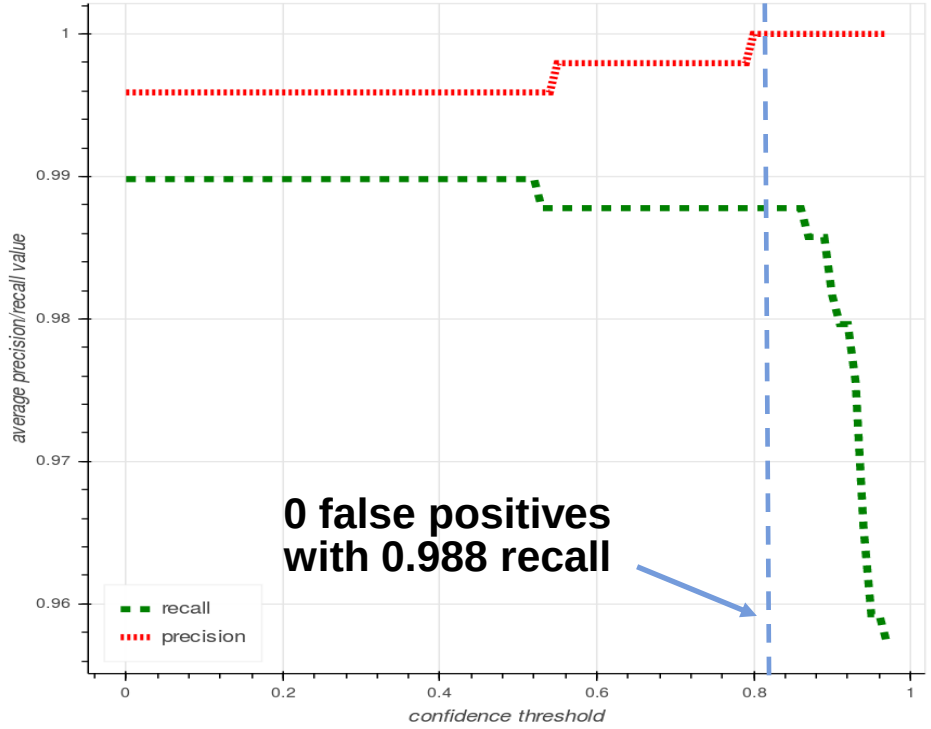
Netflix
(feature: total burst size)



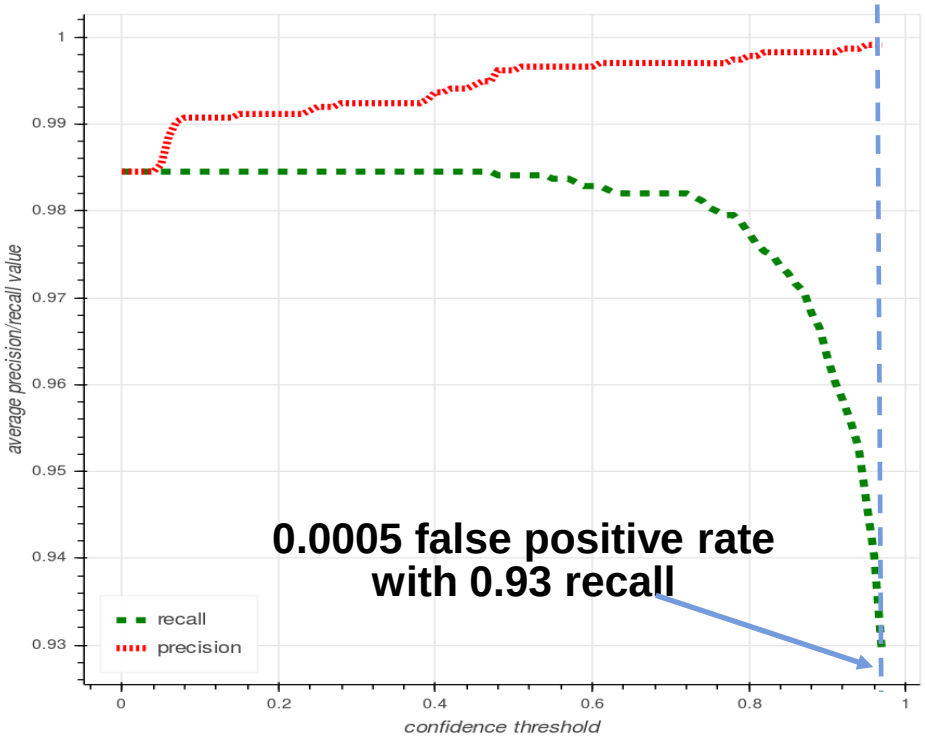
No recurrent confusions
(despite many same-series titles)

Tuning for precision

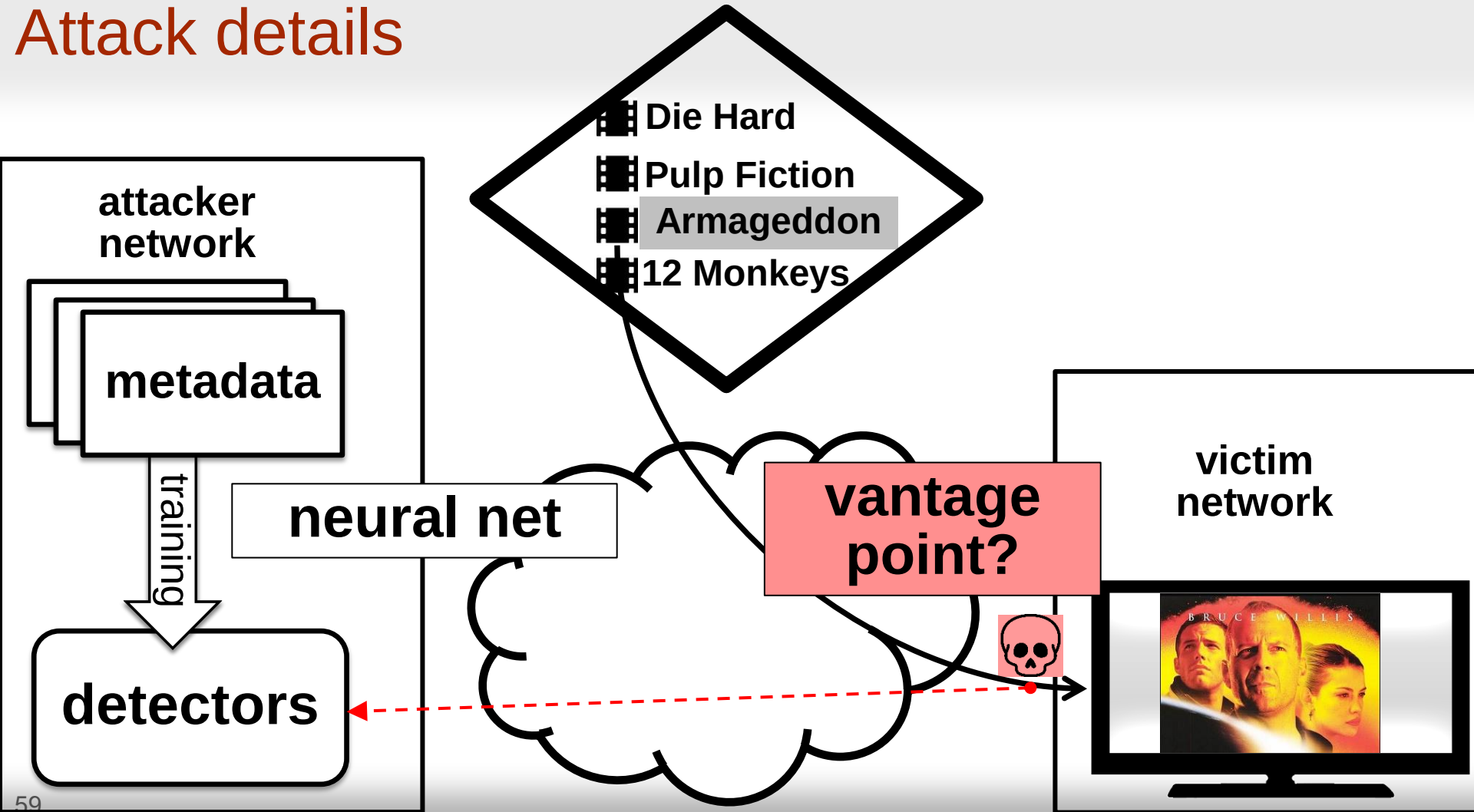
YouTube
(feature: total burst size)



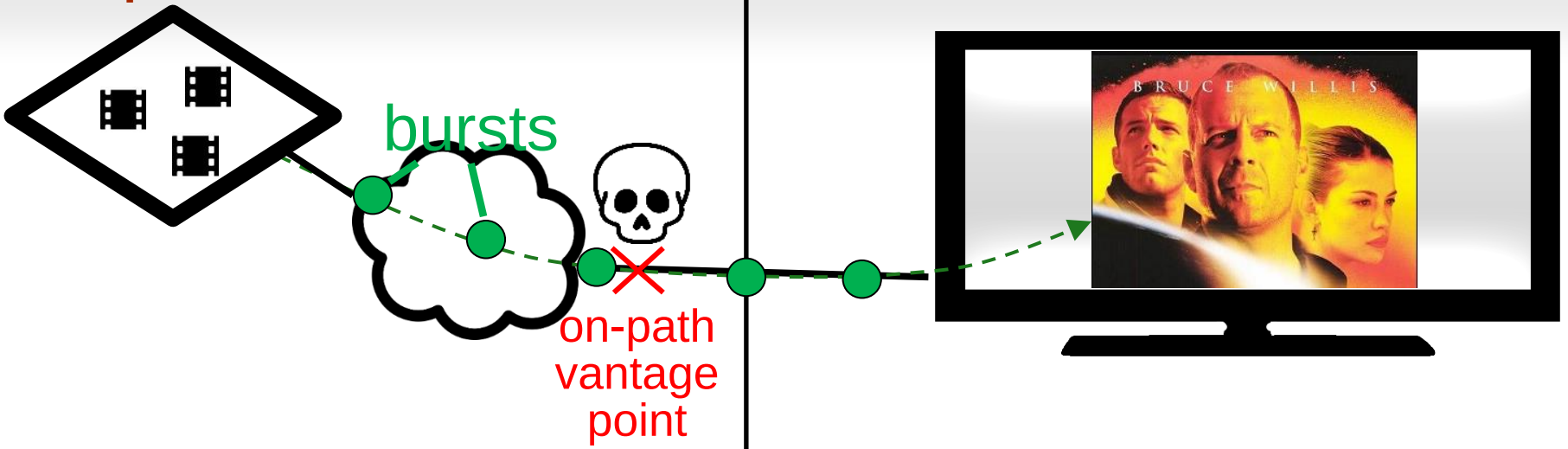
Netflix
(feature: total burst size)



Attack details

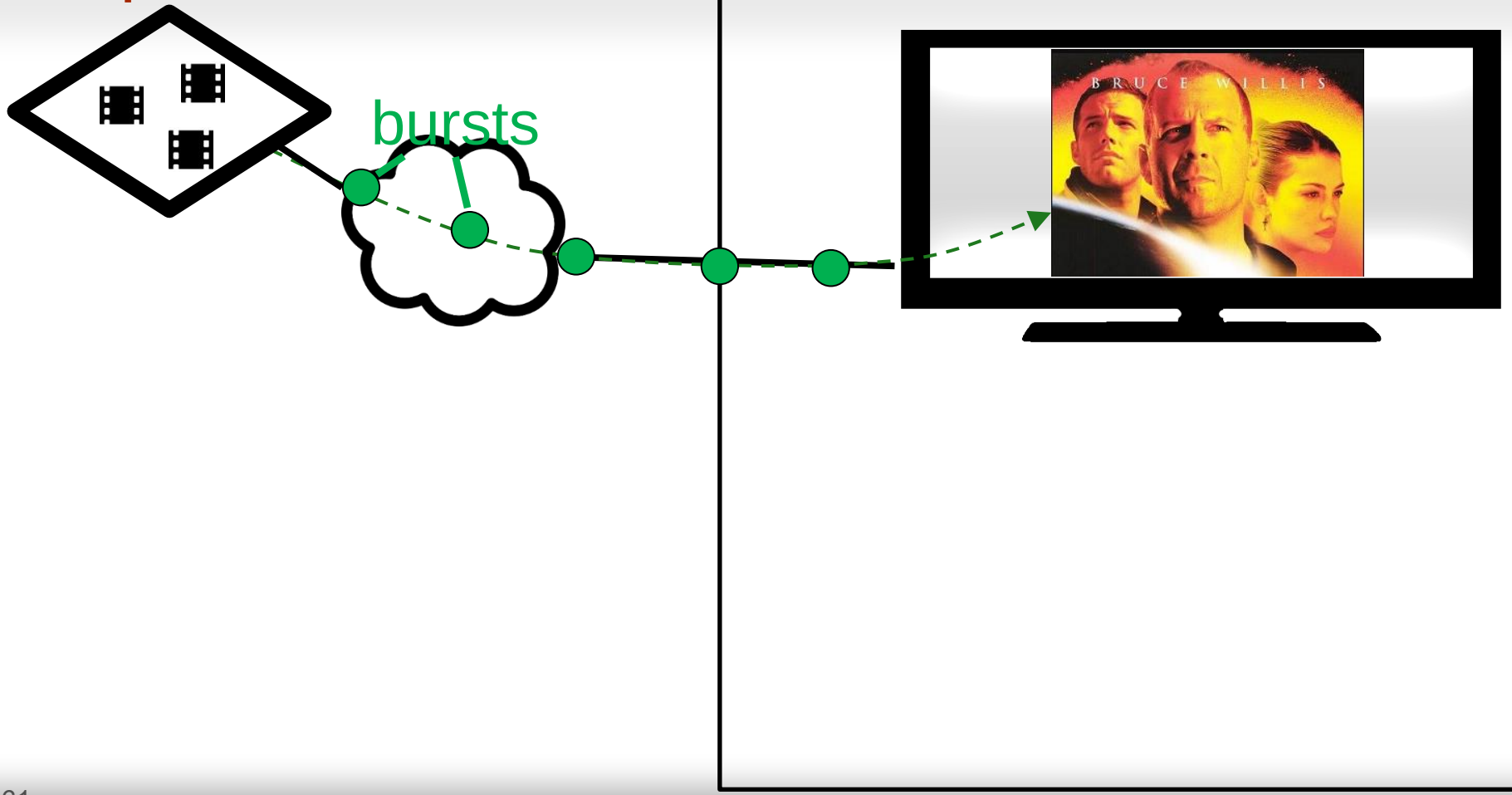


Off-path attackers

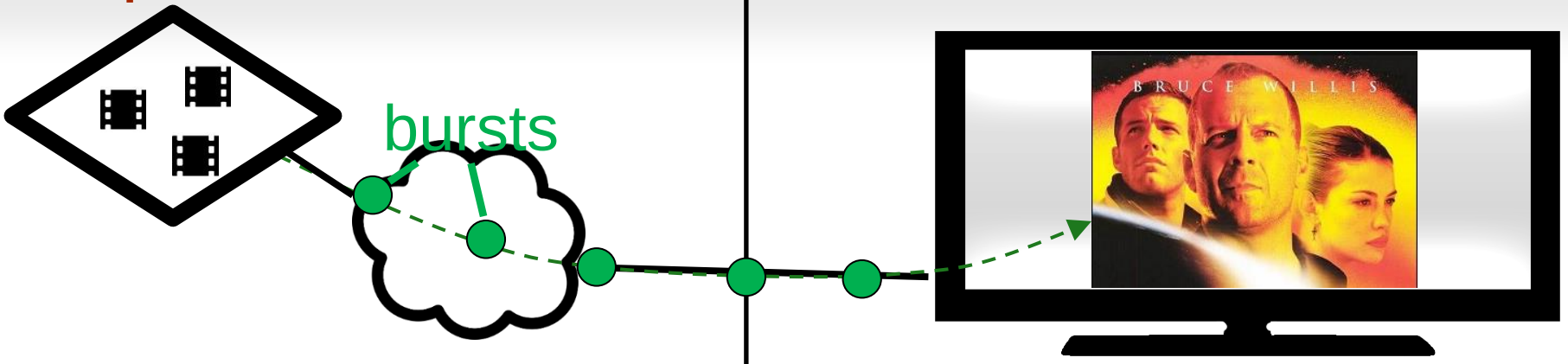


Wi-Fi access points, proxies, routers, enterprise or national network censors, ISPs

Off-path attackers



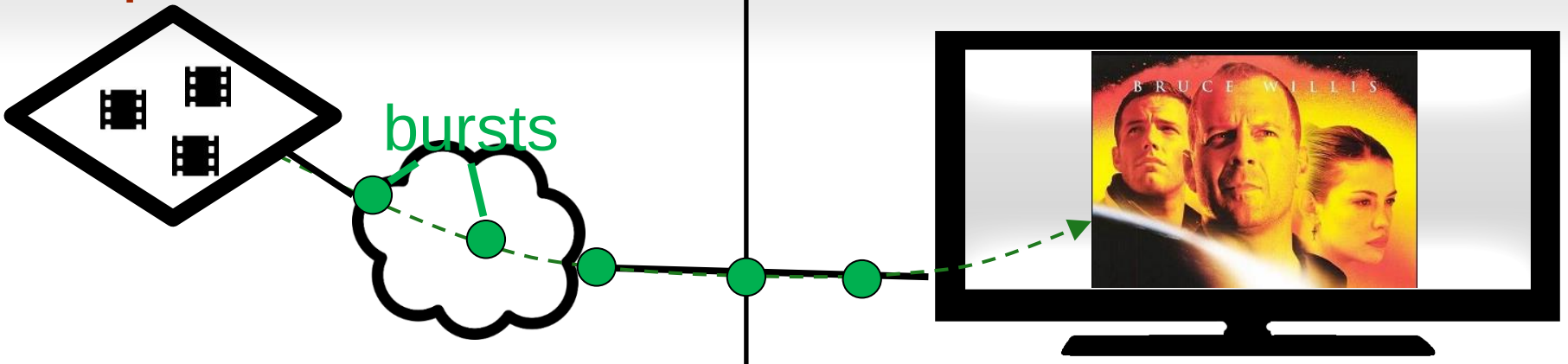
Off-path attackers



A visited webpage?

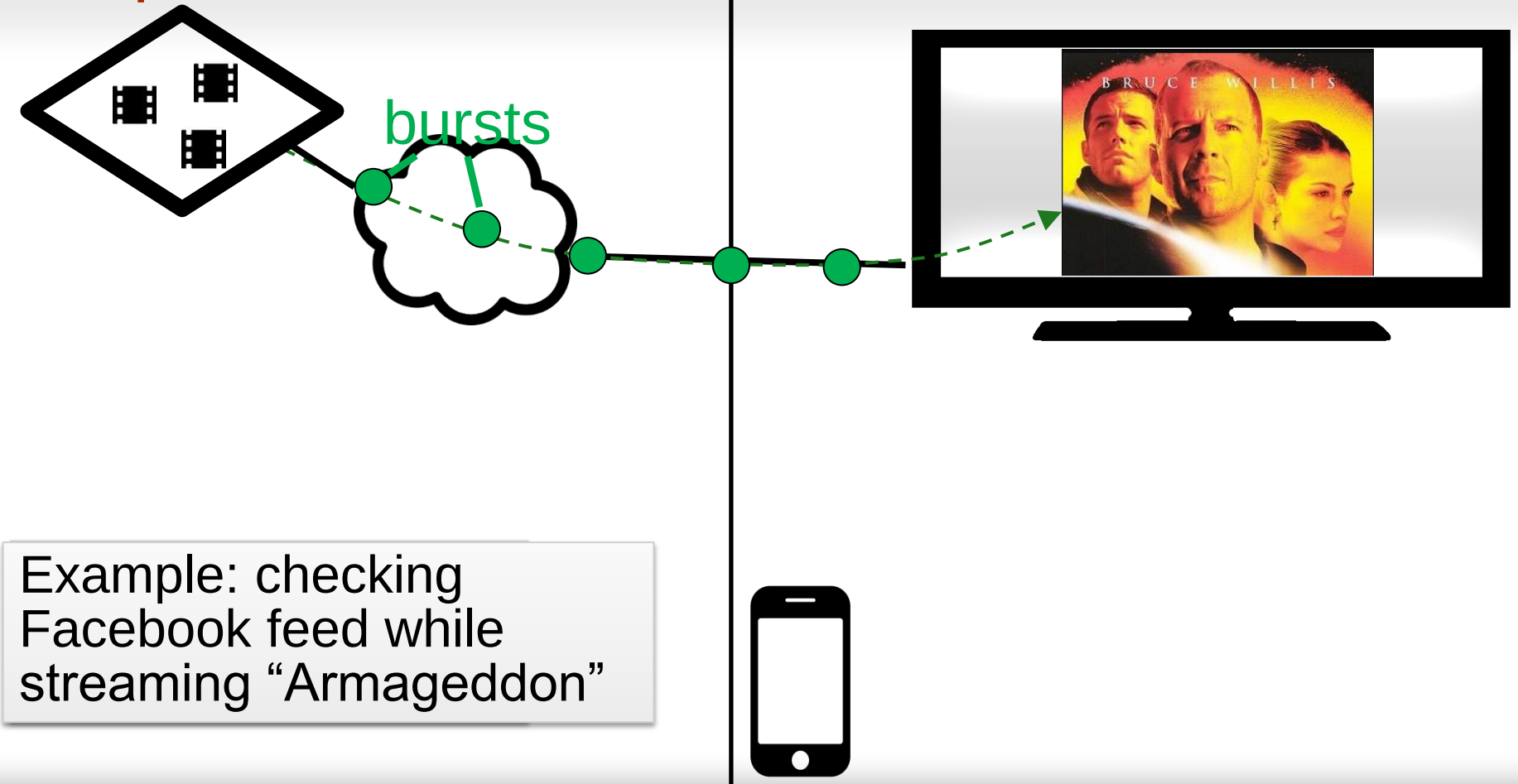
A smartphone app?

Off-path attackers



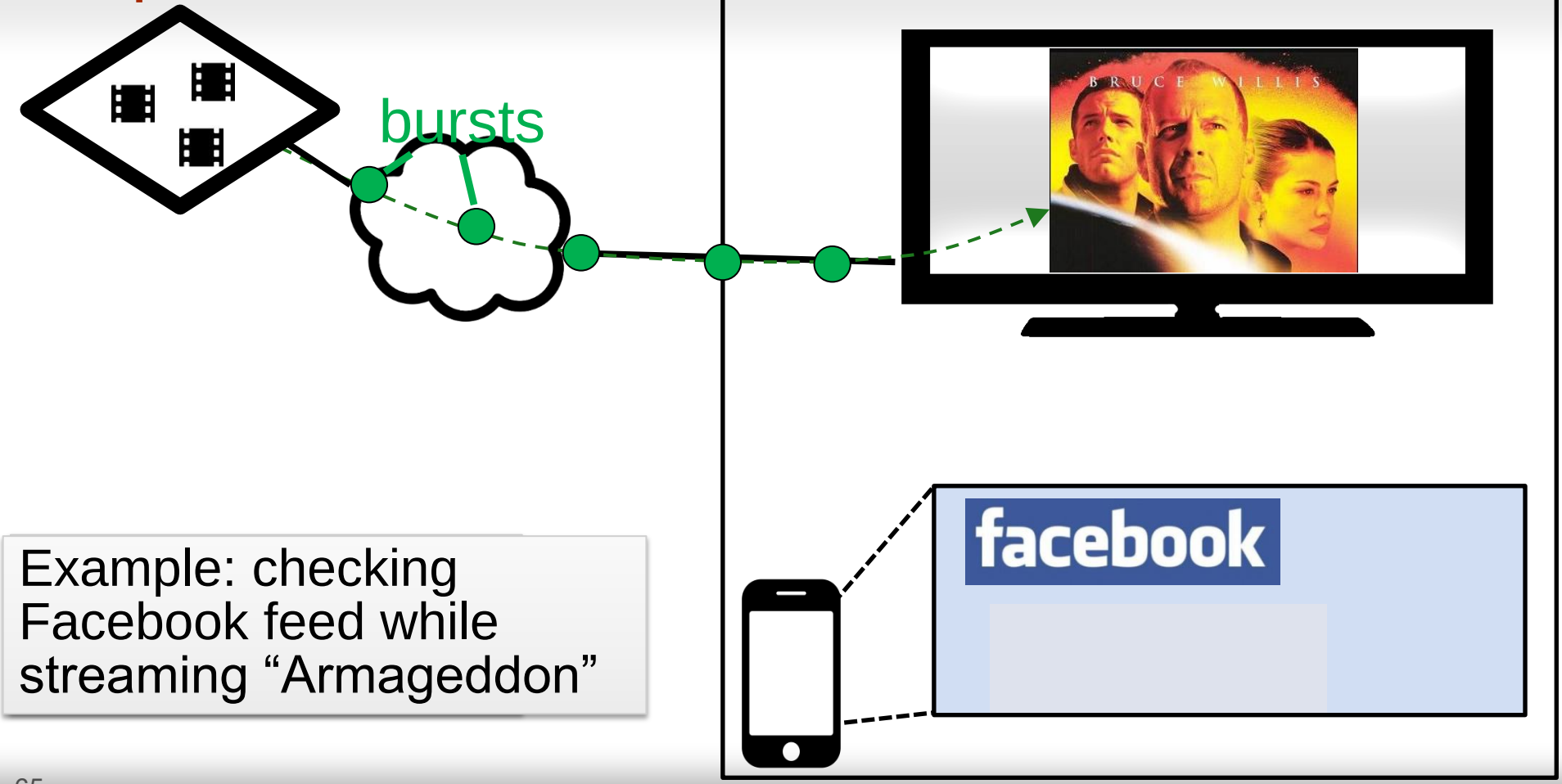
Example: checking Facebook feed while streaming “Armageddon”

Off-path attackers



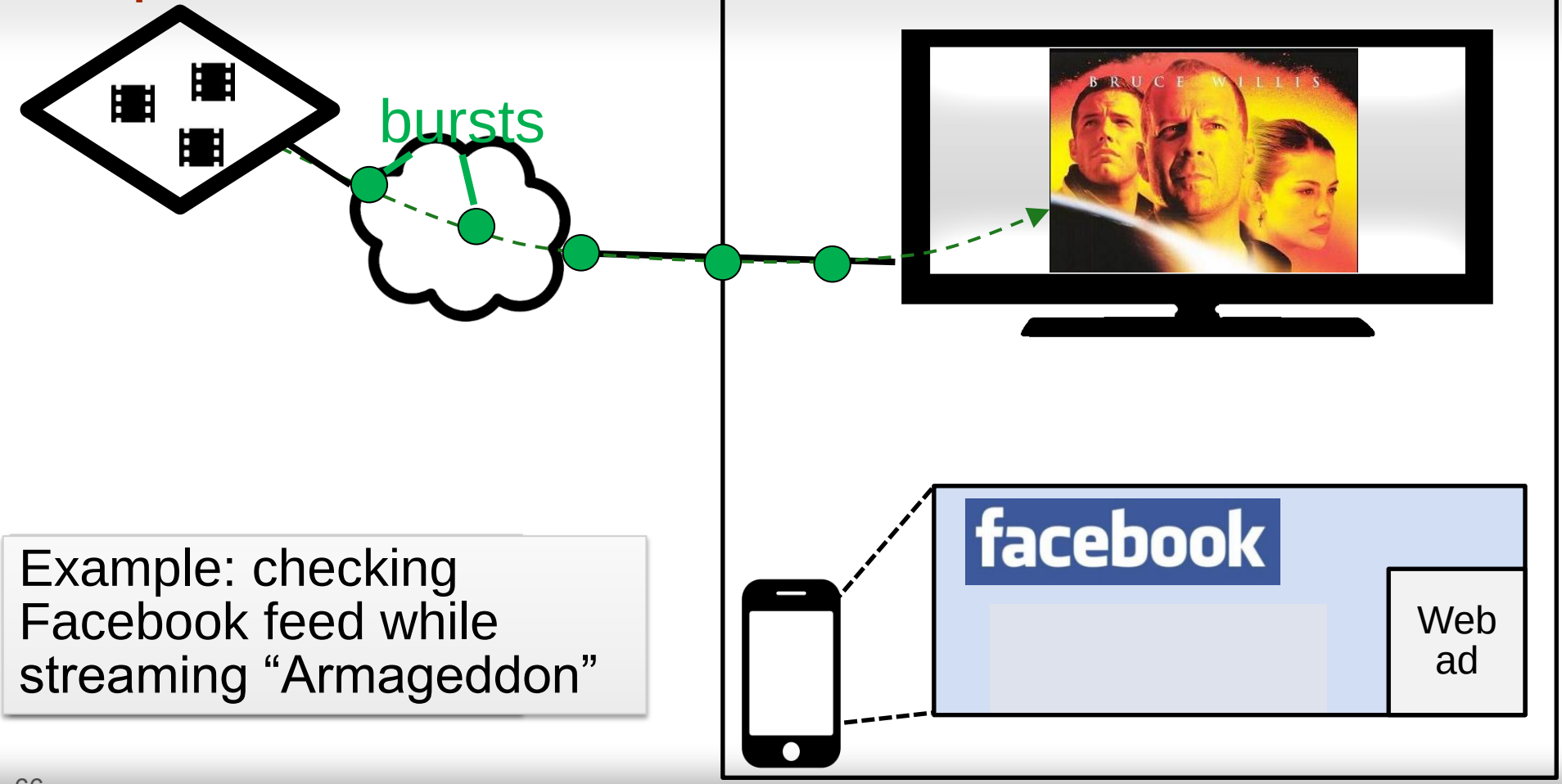
Example: checking Facebook feed while streaming “Armageddon”

Off-path attackers



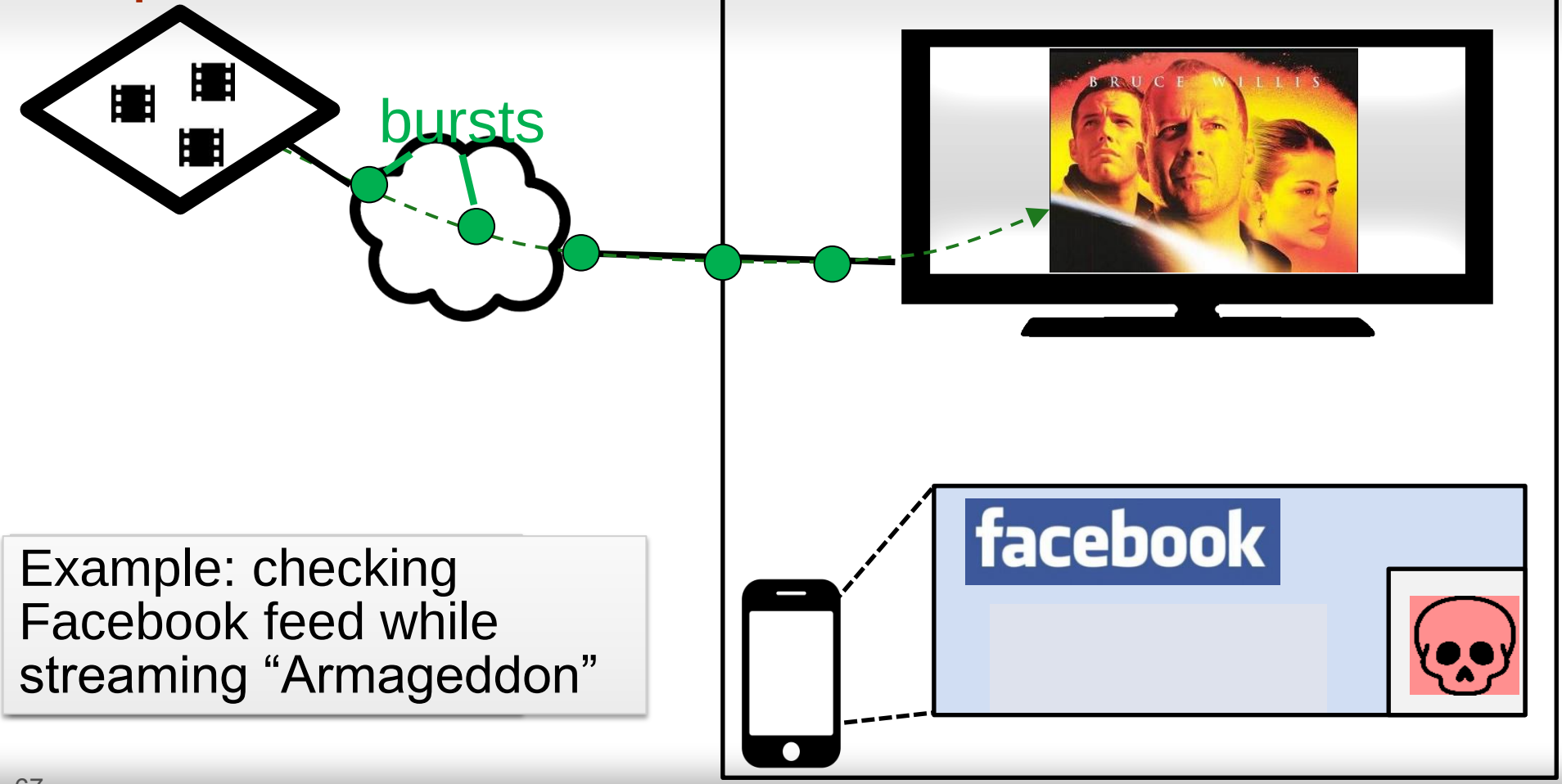
Example: checking Facebook feed while streaming “Armageddon”

Off-path attackers



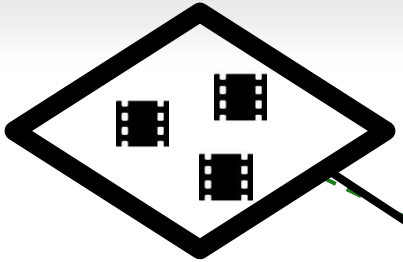
Example: checking Facebook feed while streaming “Armageddon”

Off-path attackers



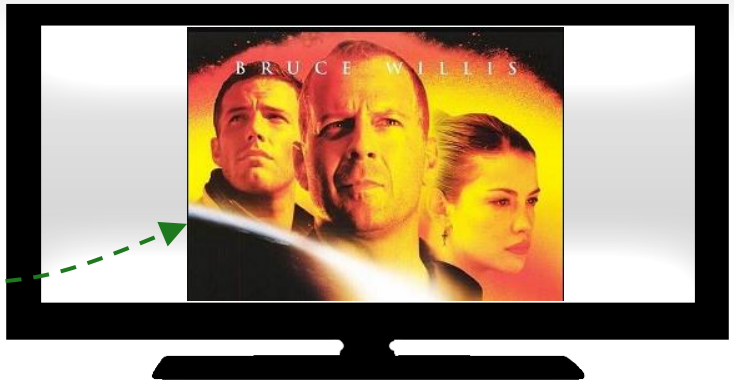
Example: checking Facebook feed while streaming “Armageddon”

Off-path attackers

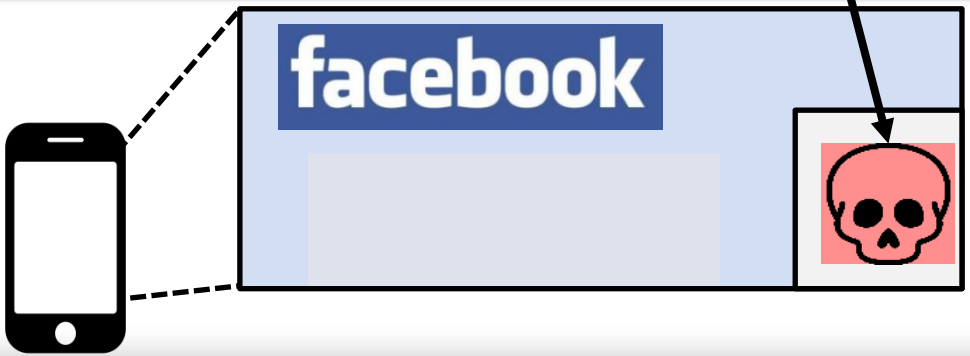


bursts

victim network

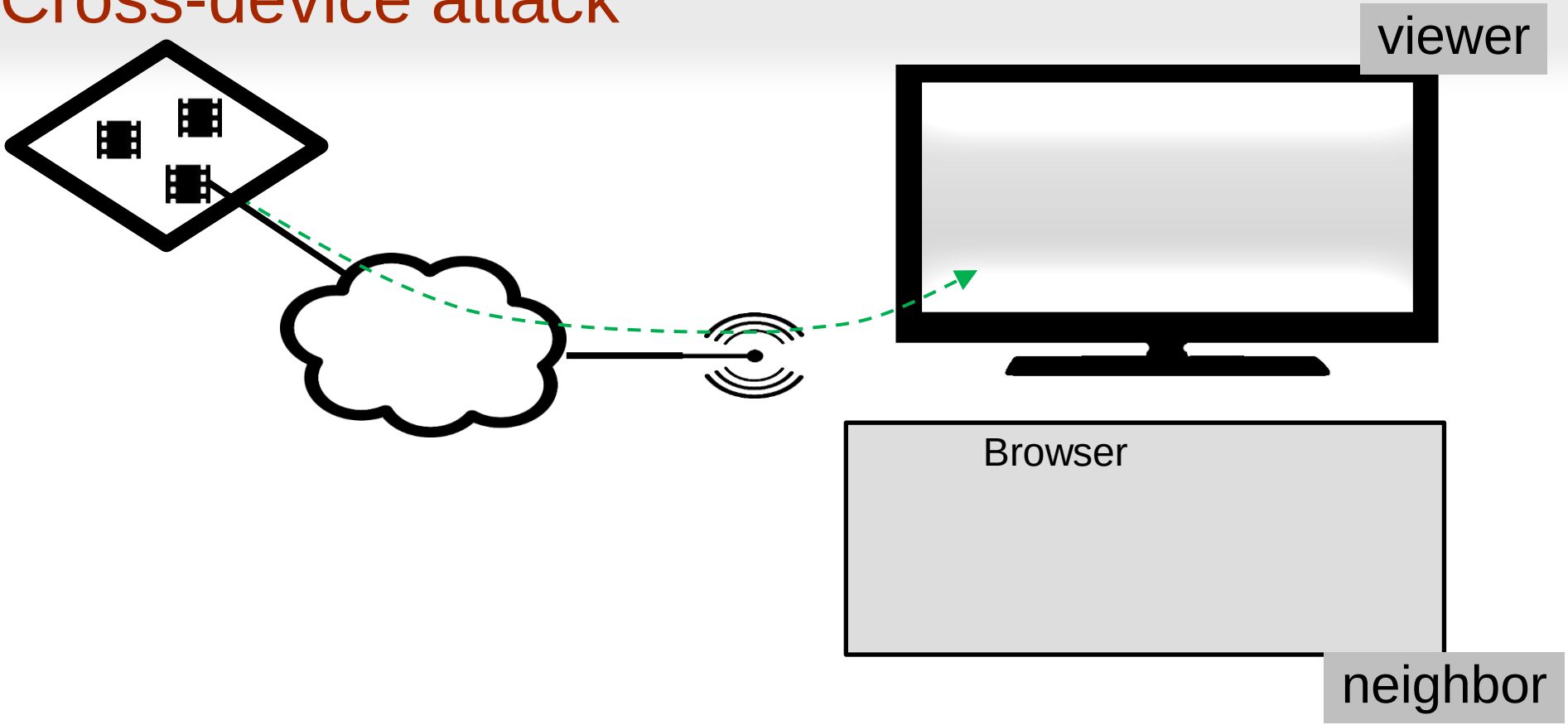


Three-fold confinement: different device, browser process, sandboxed iframe

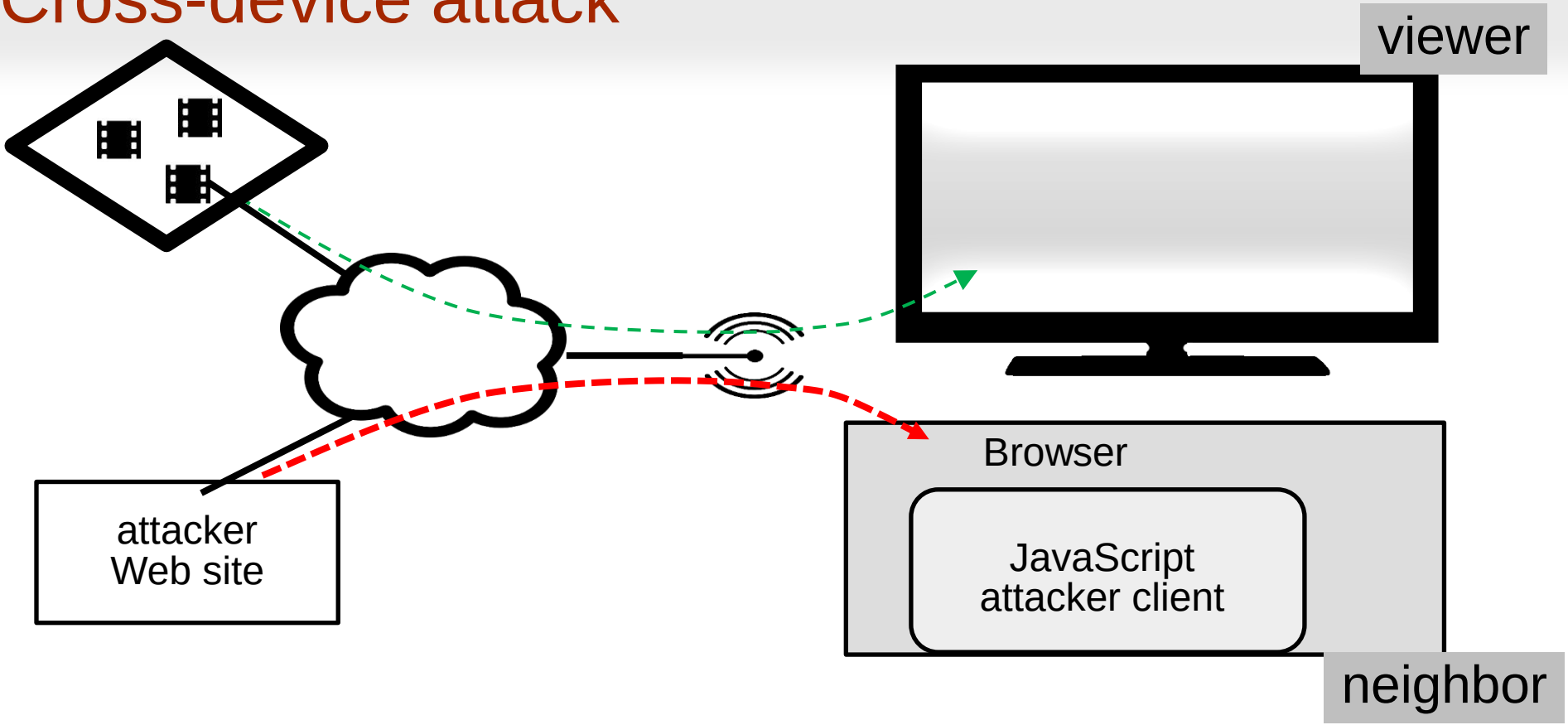


Example: checking Facebook feed while streaming “Armageddon”

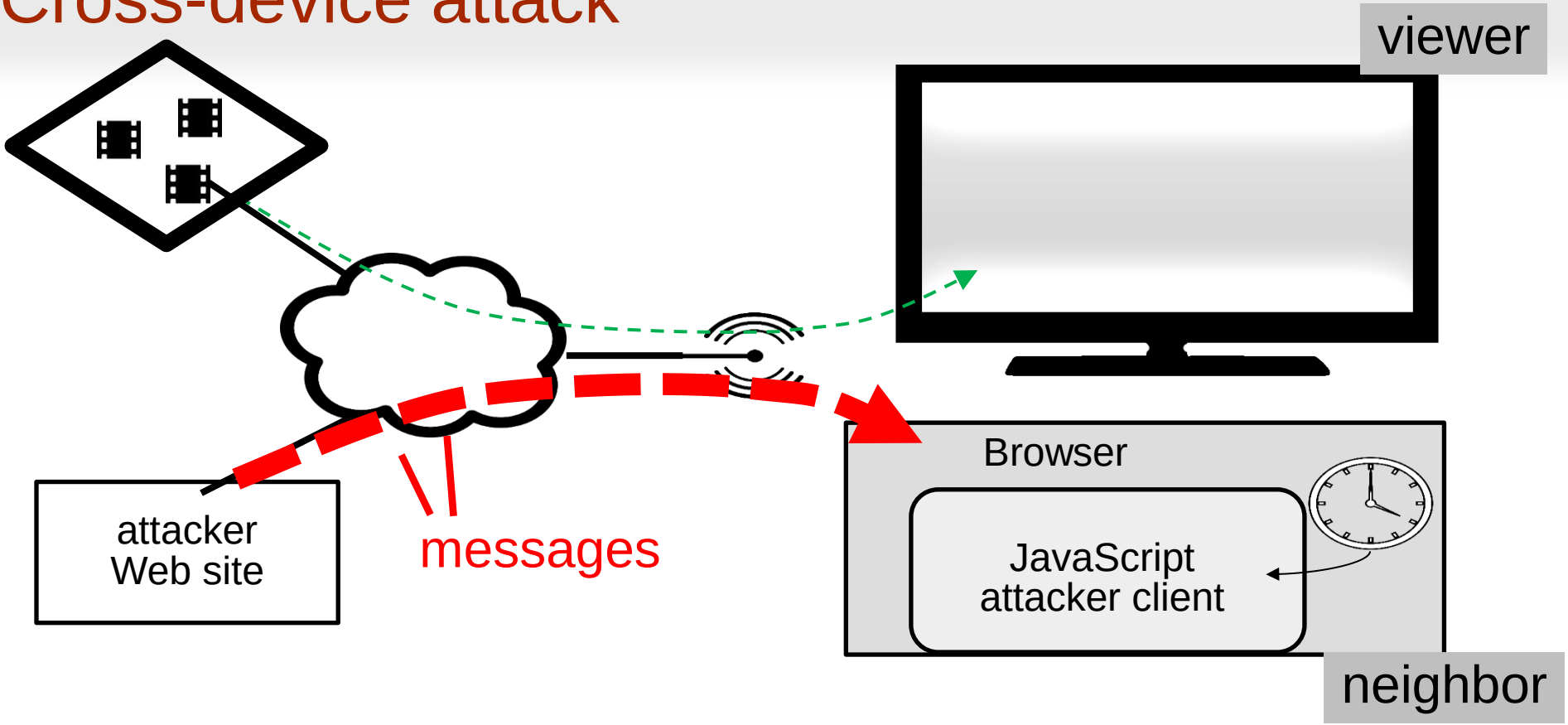
Cross-device attack



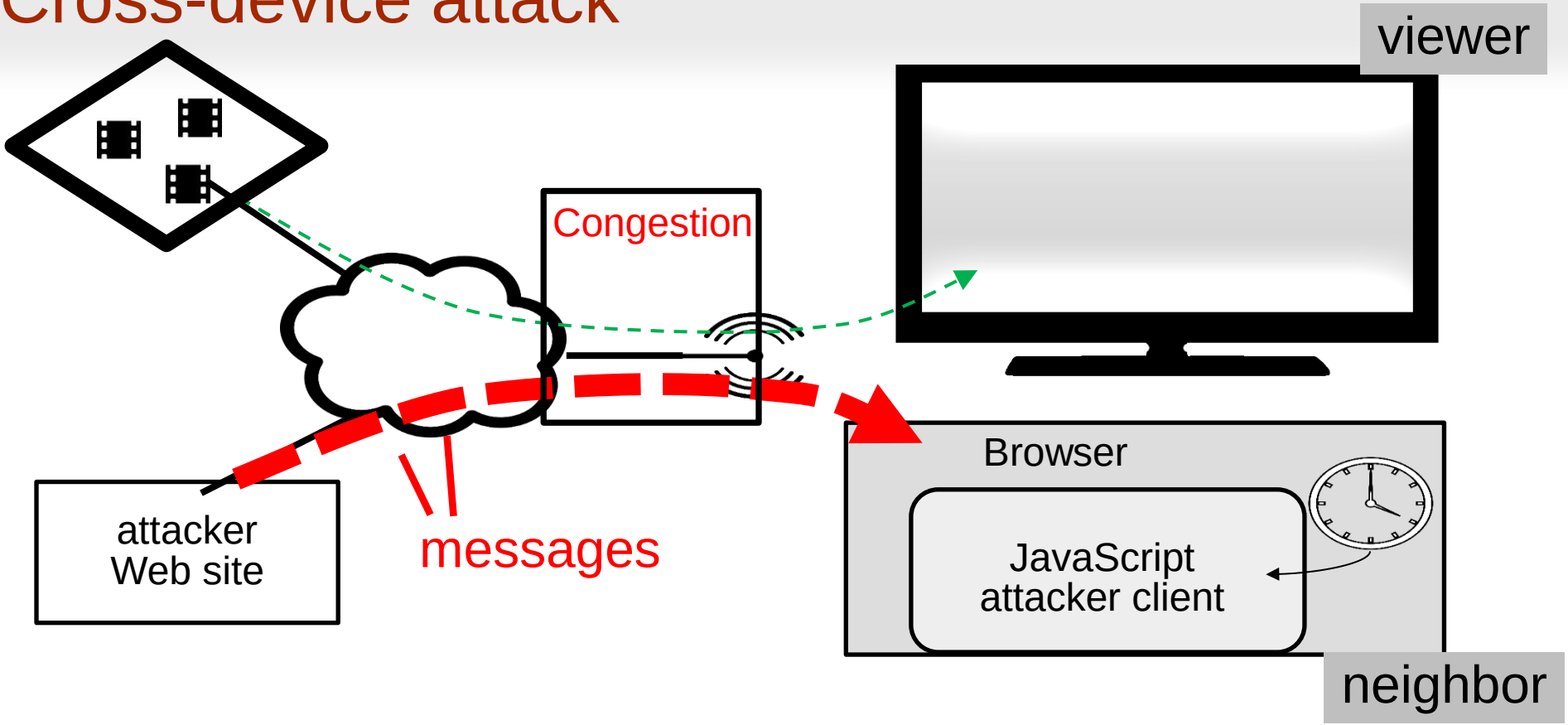
Cross-device attack



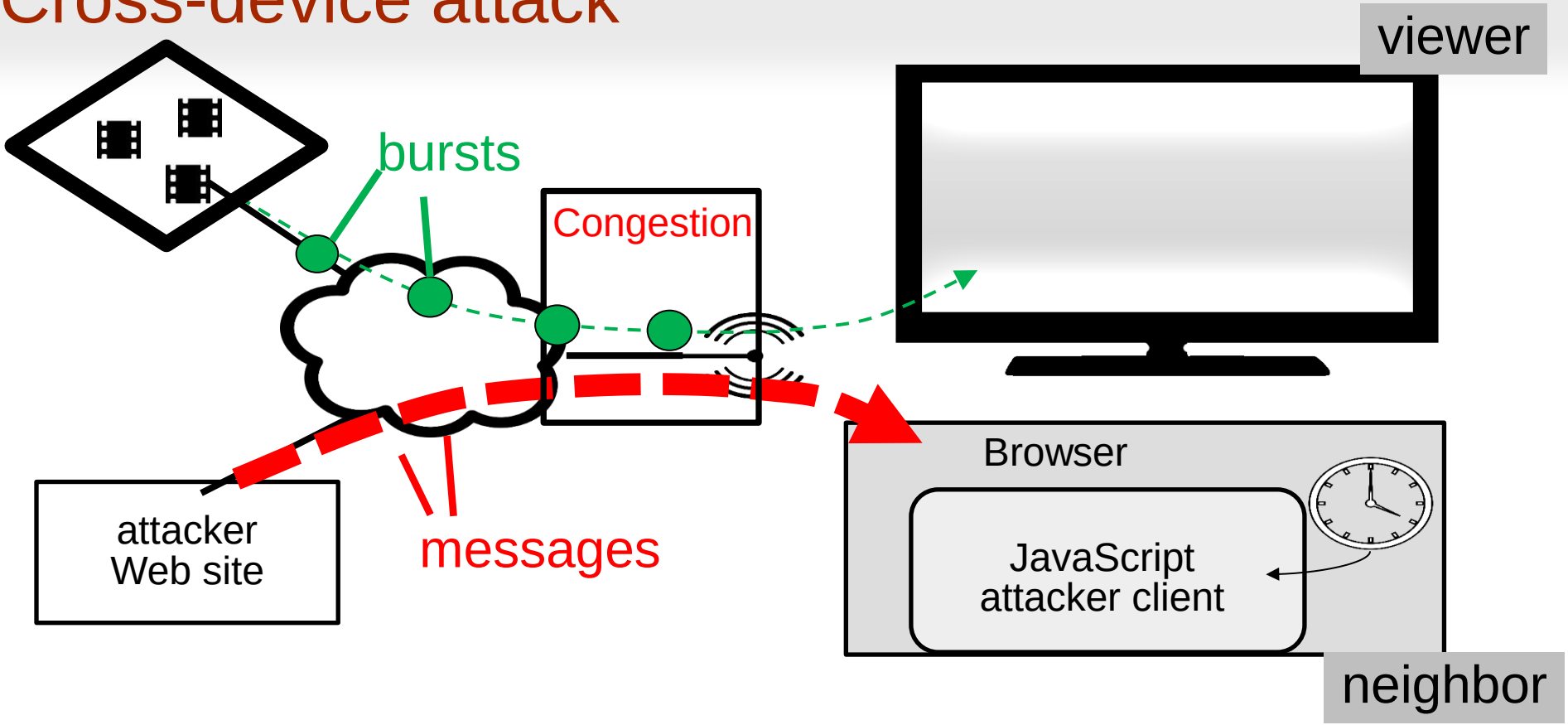
Cross-device attack



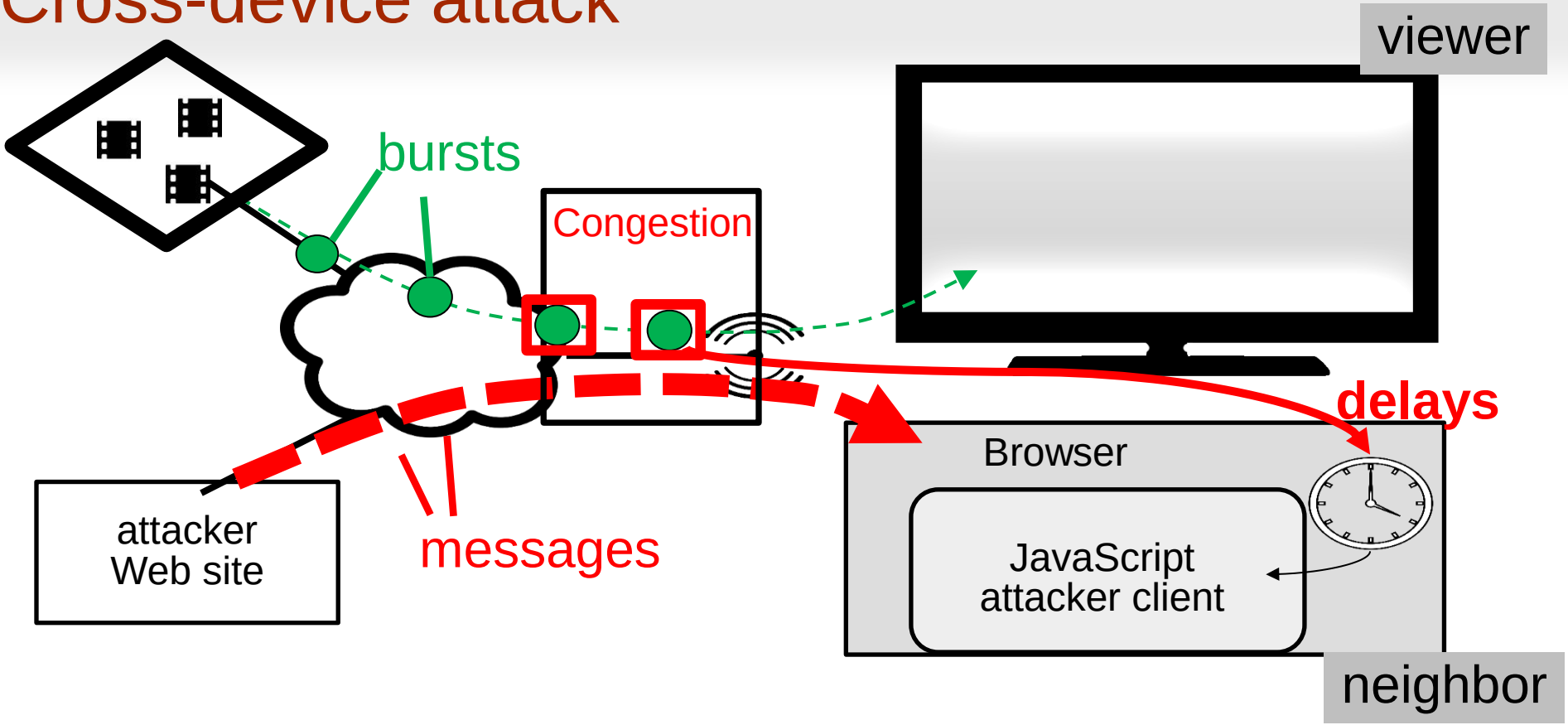
Cross-device attack



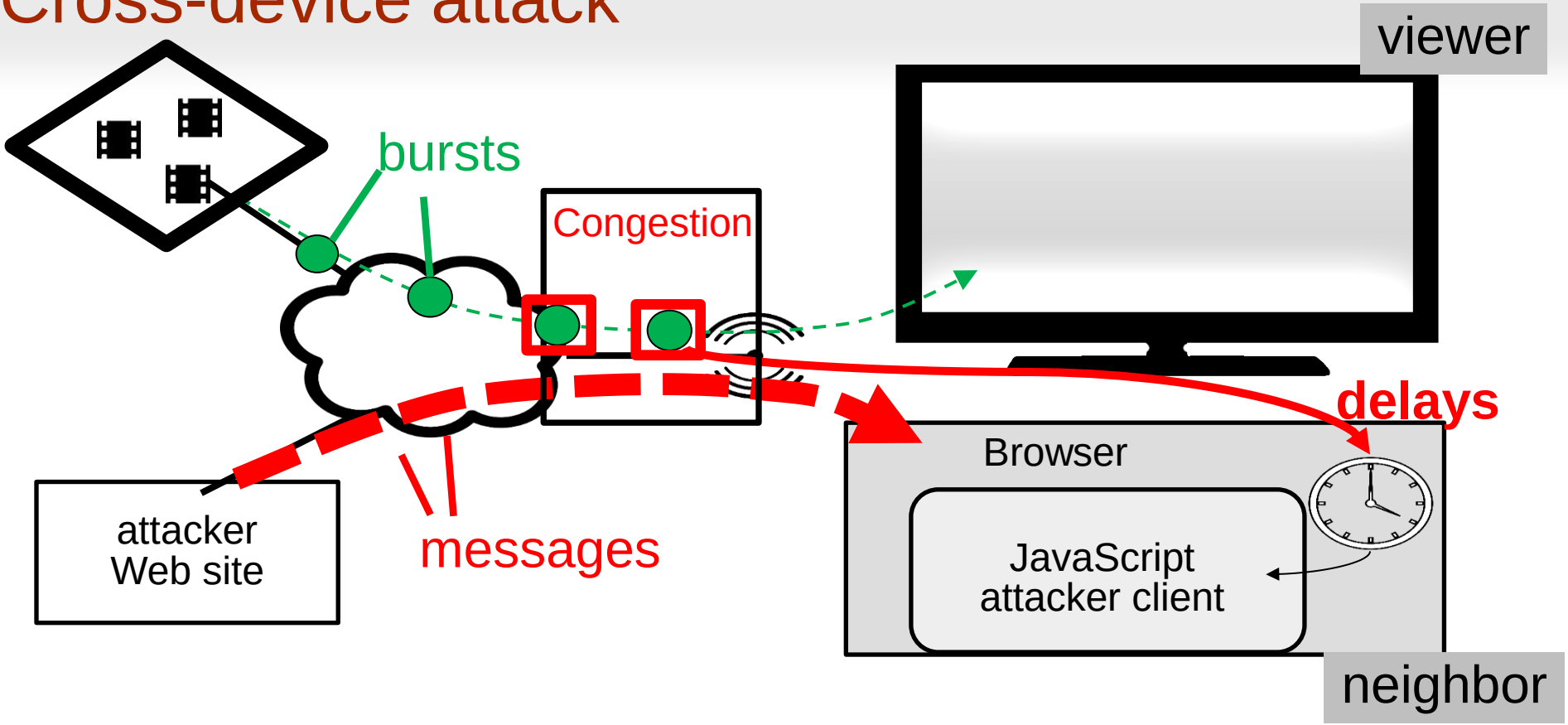
Cross-device attack



Cross-device attack

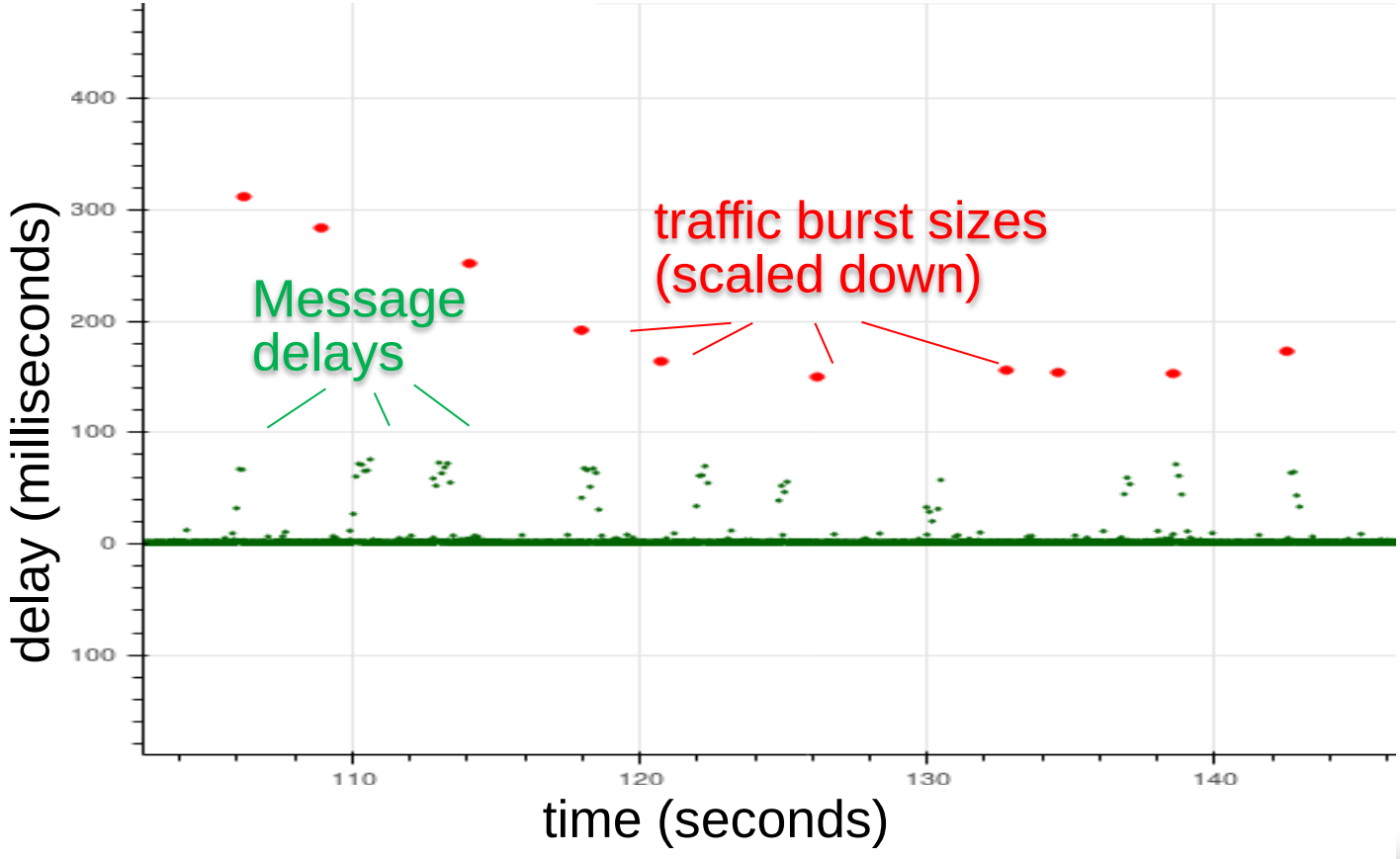


Cross-device attack

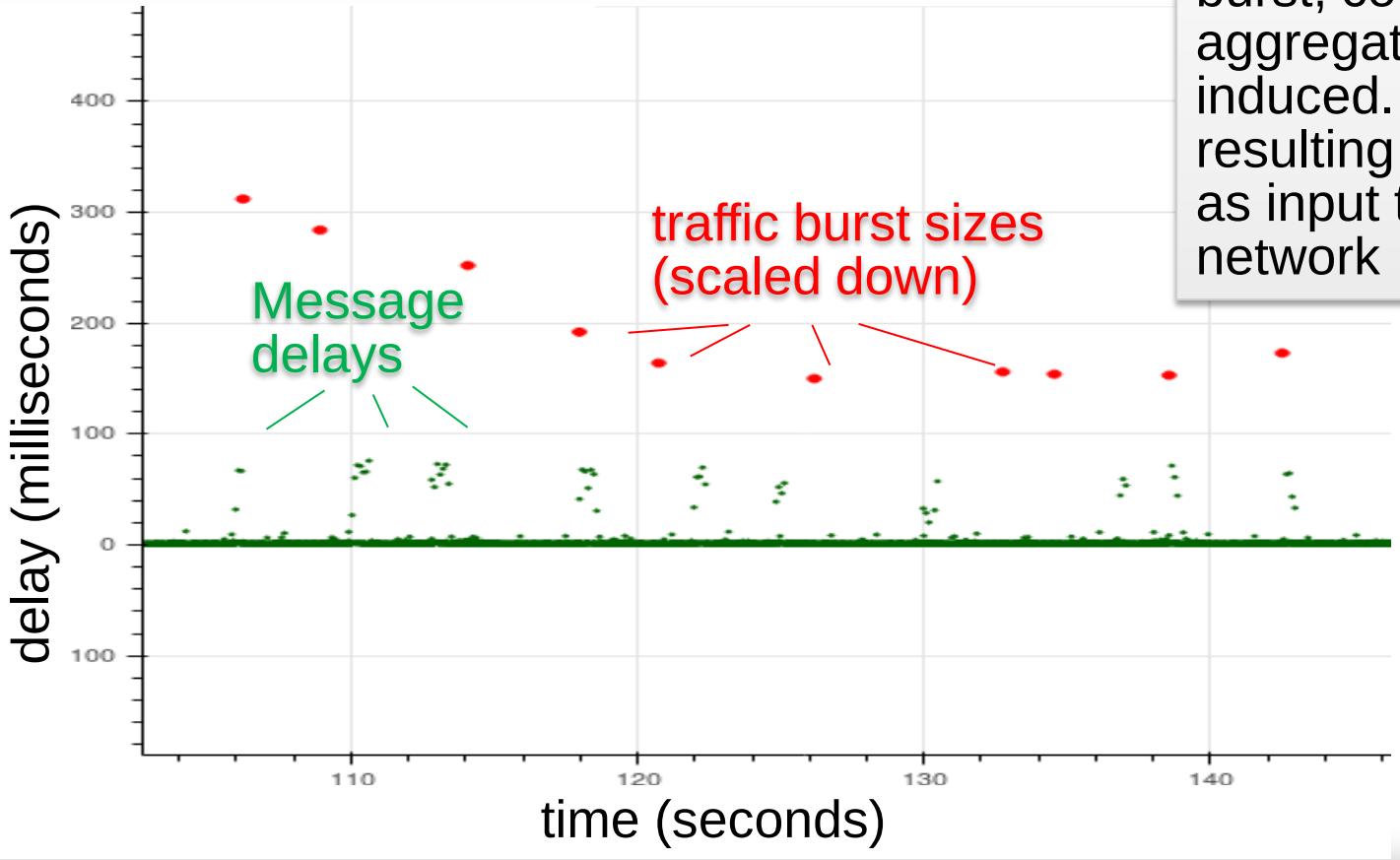


Noisy, coarse estimate of actual traffic bursts

Delay-bursts

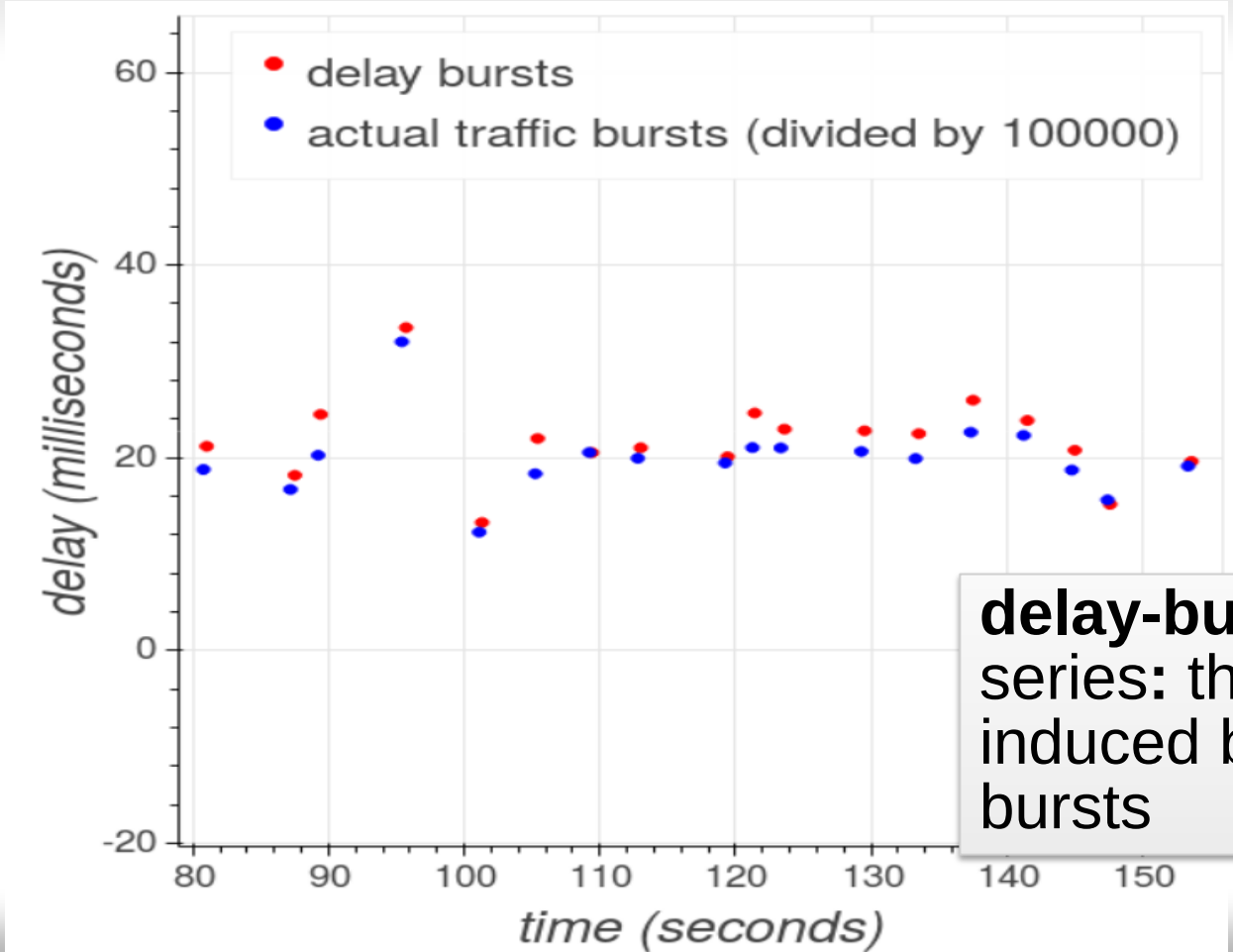


Delay-bursts



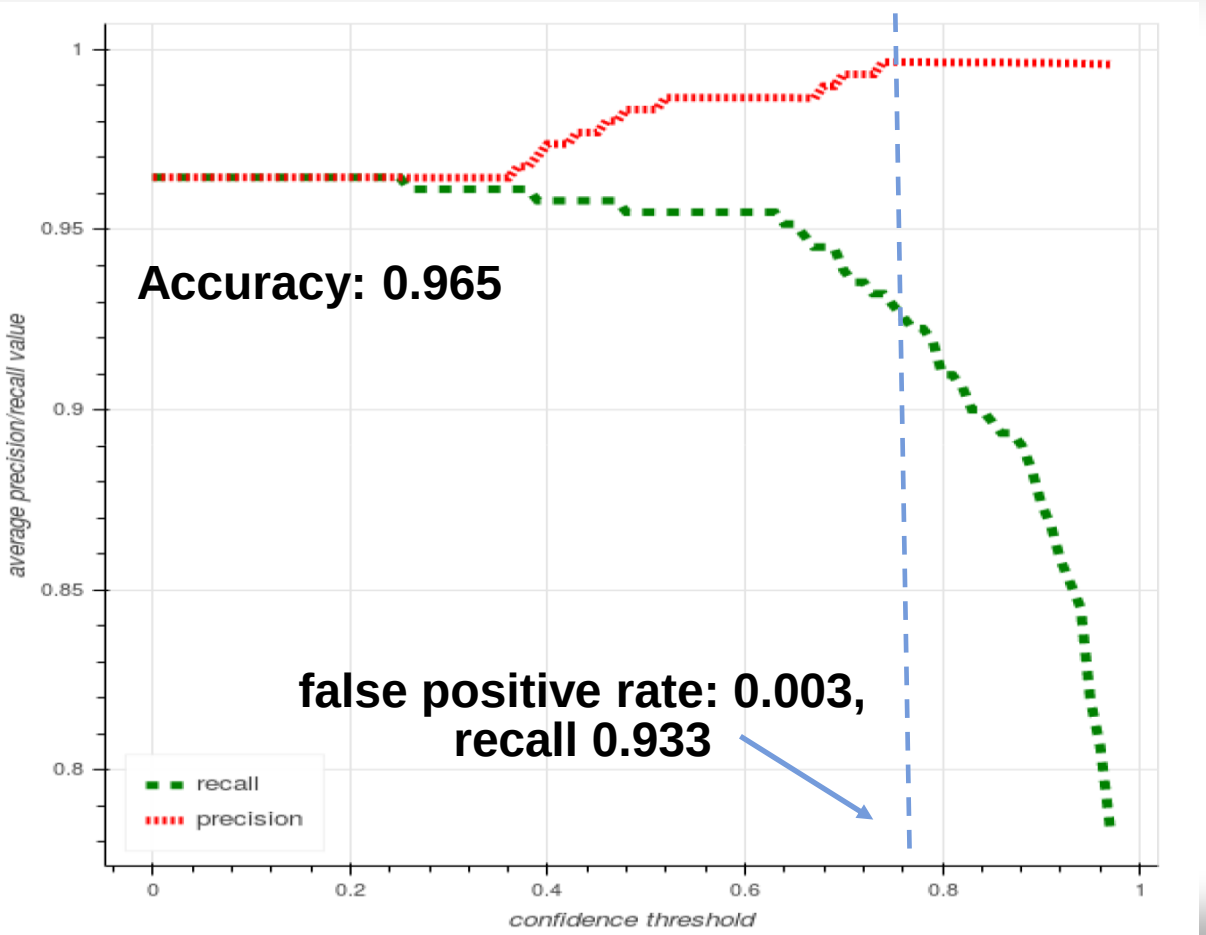
For each traffic burst, compute aggregate delay induced. Use resulting time-series as input to neural network

Delay-bursts vs. traffic bursts

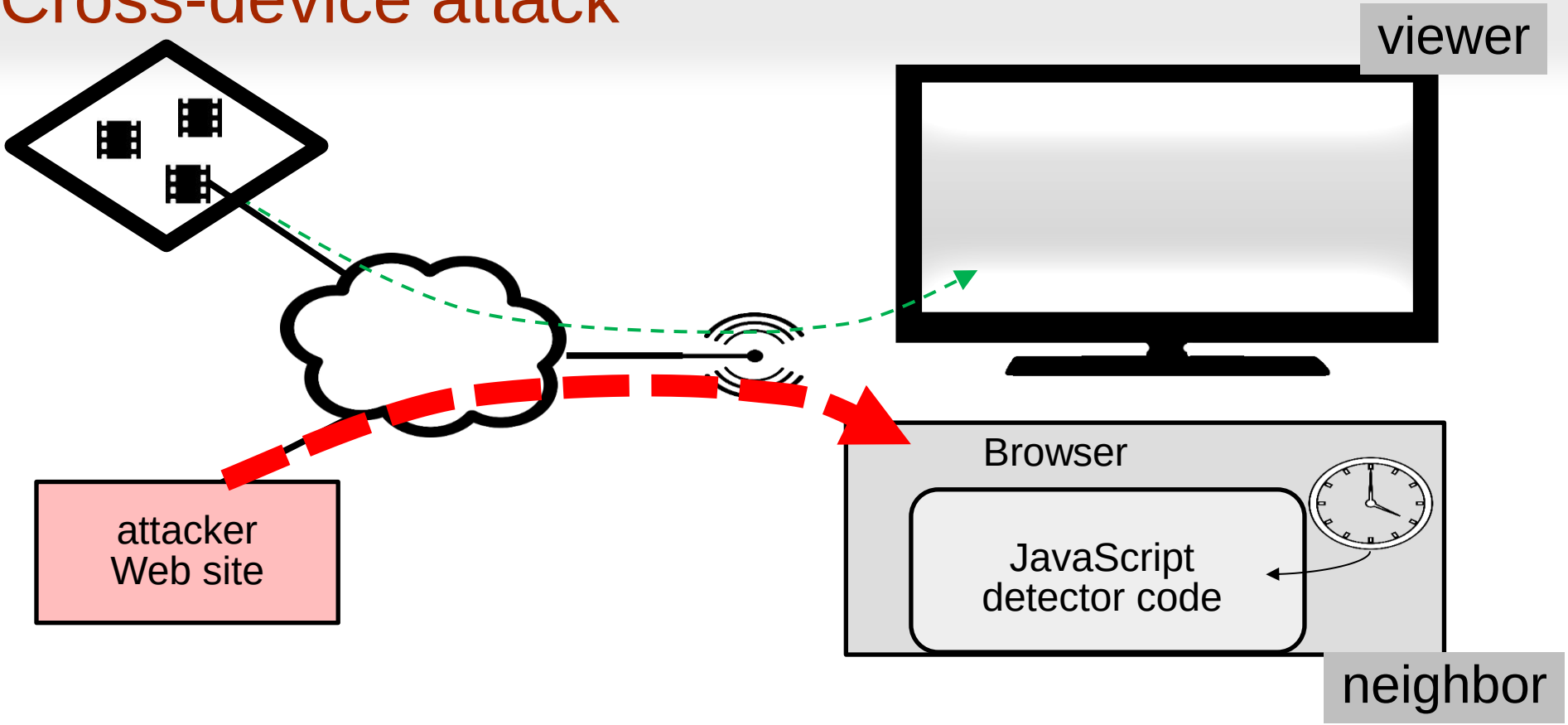


delay-bursts time series**:** the delays induced by traffic bursts

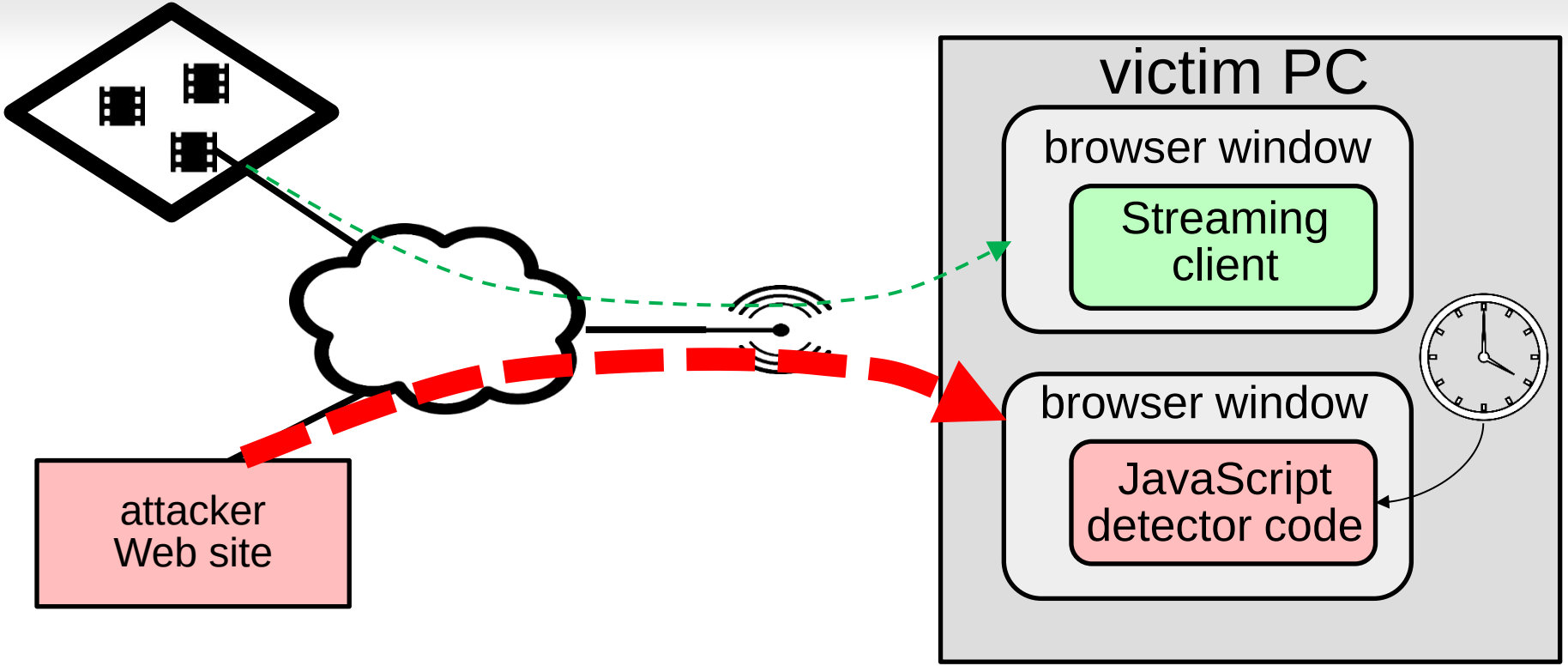
1/10 cross-device attack: precision vs. recall



Cross-device attack

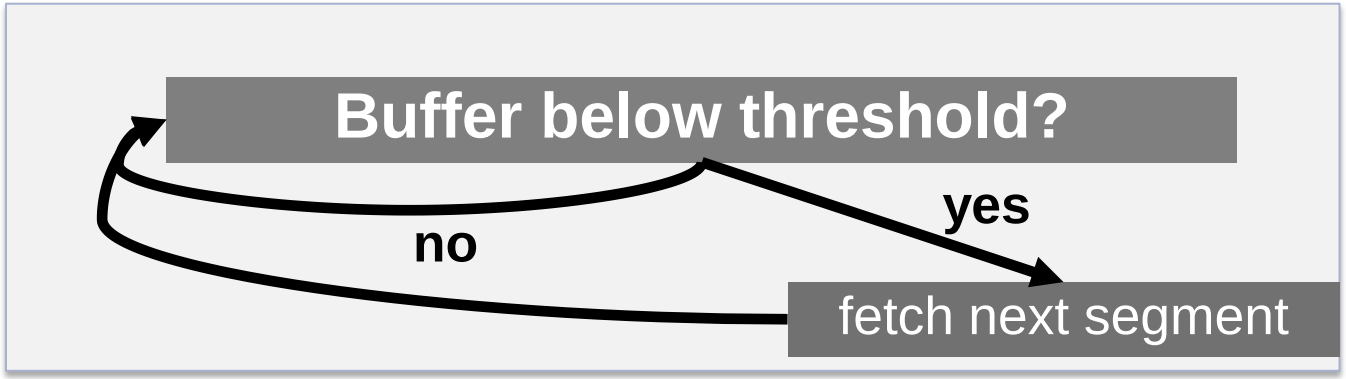


Cross-site attack



Mitigating the DASH leak

- Modern streaming traffic characteristics
  - Title bitrate pattern unique when sampled at few-seconds granularity
  - Fetching **at segment granularity** (= every few seconds)



- Maximizes “quality of experience”, server load, and network bandwidth utilization
- However, information leakage is intrinsic…

Thank you!

- Further information and the paper: https://beautyburst.github.io/