DISTRIBUTED RESEARCH ON EMERGING APPLICATIONS & MACHINES
Department of Computational & Data Sciences
Indian Institute of Science, Bangalore

DREAM:Lab

Towards Resilient Stream Processing on Clouds using Moving Target Defense

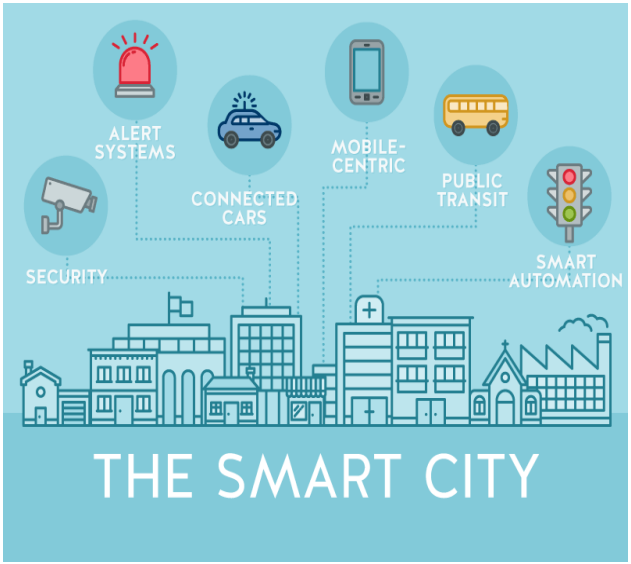
Shilpa Chaturvedi,(Member of Technical Staff, NetApp)
Yogesh Simmhan

IEEE ISORC Conference 2019

08-May-19

Motivation: Internet of Things(IoT)

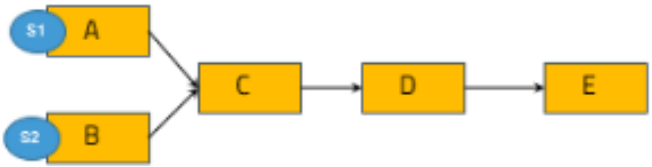
- **Billions of sensors** generating observation streams as time-series
  - *Smart Cities, Industrial IoT, Fitness devices*
- **Real-time analytics** on incoming streaming data
- **Explosion** of innovative services & apps
  - *Public services, start-ups*



http://1.bp.blogspot.com/-N5JfgIr1DaE/Vbpn0iIwYXI/AAAAAAAAjws/D1_DeUmxf8c/s1600/Smart-Cities-Features.png

Stream Applications

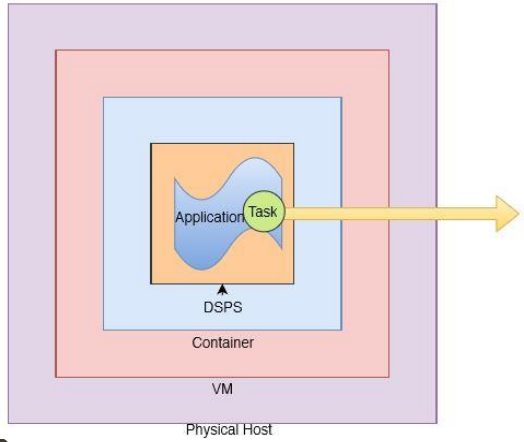
- Applications are composed as DAGs (**dataflows)**
- Set of **tasks** as vertices and set of **streams** as edges
- **Streaming data** is ingested into applications for **real time analytics**



Distributed Stream Processing System

- DSPS are **Big Data** platforms tailored for *scalable processing* of *streaming data*, with *low latency*
- Composition & distributed execution for dataflows
- Provides support for user defined task logic
- E.g. **Apache Storm, Apache Flink, Spark Streaming**

Observation



- Stream applications are deployed on Clusters
- Shared Cloud infrastructure provides scalability but also increases vulnerabilities.
- Untrusted Environment
- Multiple Layers involved in execution => more attack surfaces

Problem Statement

- Design pro-active application and platform level defense mechanisms
- Analyze performance penalties and other overheads for mechanisms
- Complement security strategies offered at the OS and other Cloud layers

Problem Formulation

Attacks in DSPS

- Attacks classification based on entry / target points
- Attack entry points : data pattern, network, task …
- Attack target points : VM, dataflow, tasks
- **Probe based attack**
  - ➢ Attackers probe system for fetching information
  - ➢ May use gained information for targeted attacks

Attacks Based on Entry Points

- Data Pattern Leaks
- False Data Attack
- VM Induced Attack
- Network Attack
- Hostile Dataflow Attack

Attacks Based on Target Points

- Data Privacy Attack
- Data Integrity Attack
- Dataflow Attack
- VM Attack
- Platform Attack

Solution Proposed

Moving Target Defense (MTD)

- Introduces spatial-temporal variations into the system
- Variations leads to information gained by attackers to become irrelevant
- Probability based model

R. Zhuang, S. A. DeLoach, and X. Ou, “Towards a theory of moving target defense in Workshop on Moving Target Defense. ACM, 2014,
C. Tunc, F. Fargo, Y. Al-Nashif, S. Hariri, and J. Hughes,

Moving Cluster Approach

- Three variants :
  - Vary the mapping of tasks to VMs
  - Vary the underlying VMs in Cluster
  - Hybrid of both
- Benefits : Decreases chances of tasks being compromised
- Cost : Redeployment of tasks / tasks migrations/ additional VMs
- Protects against: VM targeted attacks, VM Induced attacks

XOR Event Payload

- Data encryption is common for data protection
- Simple bitwise XOR of event payload
- Low compute cost
- Benefits : Data is not directly readable
- Cost : XOR operation on payload / Periodic mask updates might require pausing task
- Protects against : Data Privacy Attack

Random Broker Redirections

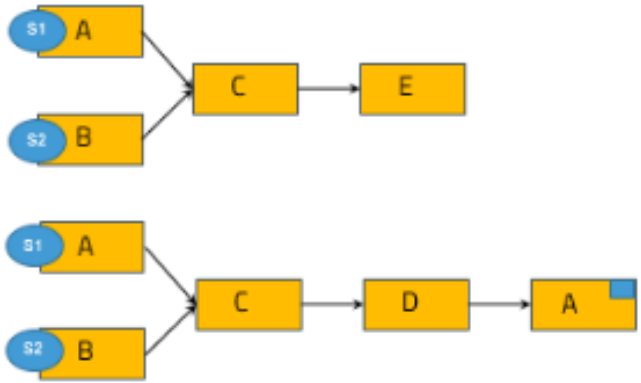
- Introduces a third-party event broker redirect event streams between upstream and downstream tasks
- Benefits : Masks the connectivity between tasks in the dataflow and DAG structure
- Cost : Control Signals, Latency
- Protects against : Data Pattern Attacks

Varying Dataflow Structure

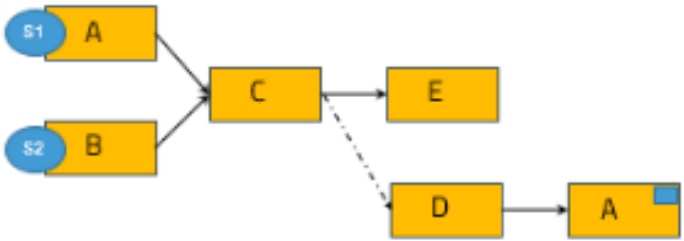
- Introduce dummy task(s) between two adjacent tasks in the DAG.
- Introduce dummy events into streams
- Benefits : Conceals DAG structure
- Cost : Additional tasks instances / workers deployment
- Protects against : Dataflow targeted attack, Data Pattern attack

Stream Reuse

- Earlier work [1] replaces overlapping dataflows with a single merged dataflow to avoid redundant



Already deployed dataflow



Possible reuse of derived stream from C

[1] S. Chaturvedi, S. Tyagi, and Y. Simmhan, "Collaborative reuse of streaming dataflows in iot applications," IEEE eScience Conference, 2017.

Stream Reuse (Contd.)

- Benefits : Modifies dataflow structure, resource savings
- Cost : DAG redeployment, Control Signals
- Protects against: Dataflow Targeted Attack

Varying Execution Units

- Varying the system configuration such as ports being used, number of workers, changing IP addresses
- Benefits : Prevents attacks exploiting configuration settings
- Cost : Redeployment of dataflows
- Protects against: Network Attack

N- Versions

- Executing multiple (n) functionally equivalent copies of an application
- Similar to ***decoy*** systems.
- Allows Voting / majority approach
- Benefits : Voting approach helps in case of corrupted tasks execution
- Cost : Multiple tasks instances running
- Protects against: Dataflow Targeted Attack

Evaluation

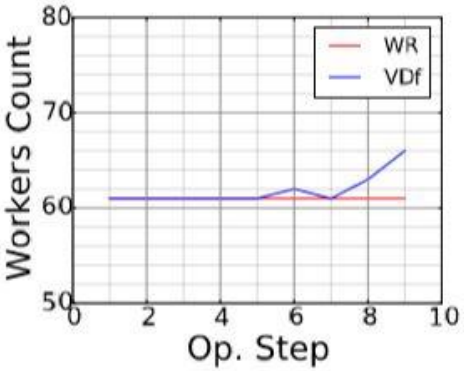
Experiment Setup

- Dataflows from Open Provenance Models for Workflows (OPMW) public repository
- Apache Storm v1.0.2, JRE v1.8
- Cluster with 8 nodes each having an AMD Opteron 3380 8-core CPU@2.6 GHz, 32 GB RAM, a 256 GB SSD, and GigaBit Ethernet, running CentOS v7

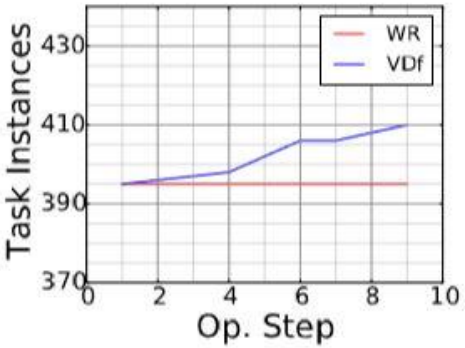
Preliminary Results

WR => Without resilience
Vdf => Varying df approach

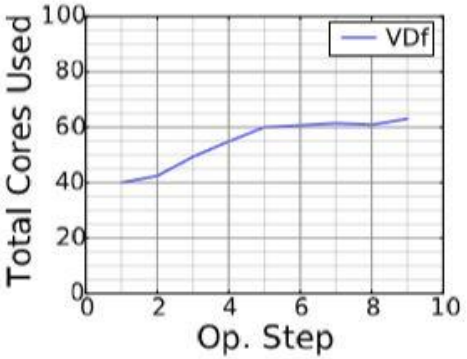
- Varying Dataflow Approach



(a) Workers Count



(b) Tasks Instances



(c) Cumulative Cores

Tasks instances increases by 3.8%

Conclusion

- Examined different attack entry / target points for DSPS
- Extended 7 MTD based approaches
- Proposed implementation and validation for the Apache Storm DSPS

Future Work

- We plan to empirically validate all the proposed strategies on a DSPS platform
- We plan to explore non MTD based approaches for providing resilience
- Exploring latency guarantees along with reuse

DISTRIBUTED RESEARCH ON EMERGING APPLICATIONS & MACHINES
Department of Computational & Data Sciences
Indian Institute of Science, Bangalore

Thank You!

info.shilpac@gmail.com