Using a Distributed Object-Oriented Database Management System in Support of a High-speed Network Intrusion Detection System Data Repository

2Lt Phillip W. Polk, GCS-01M

The Problem

- Current cutting-edge Intrusion Detection Systems - in particular, those of the USAF - use relational database management systems (RDBMS) for long-term incident storage and correlation engine repositories

Implications

- Network intrusion events are naturally represented as objects, BUT
- Relational databases are not well-suited for distributed systems and the rapid capture of object-form data from intrusion detection sensors

Research Hypothesis

- A well-balanced distributed object-oriented database management system (DOODBMS) more naturally stores network incident data, resulting in significantly higher data storage rates without substantially increasing data retrieval overhead.

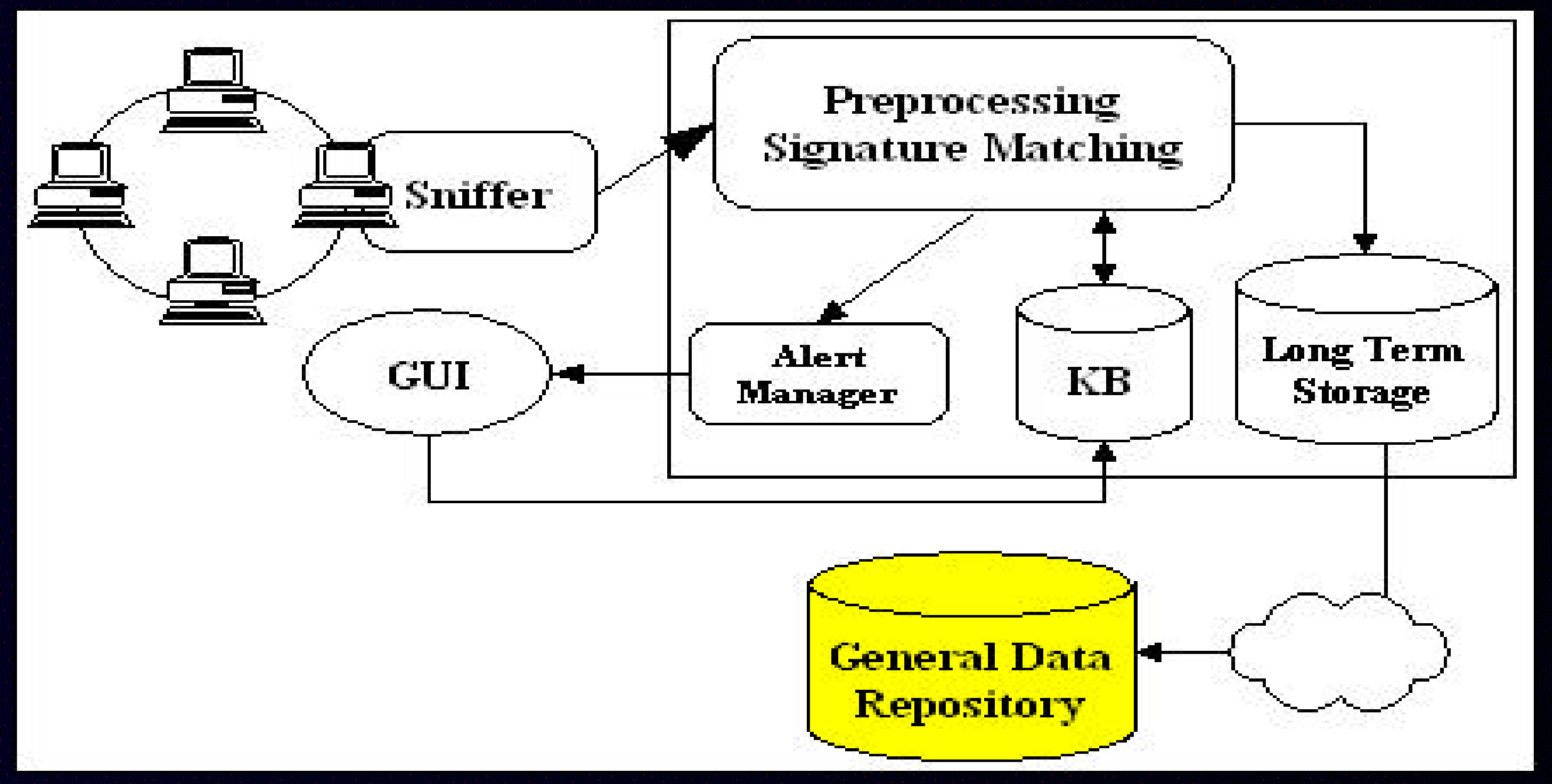
In English: Replacing the RDBMS with a DOOBMS can remove the storage bottleneck without causing a retrieval bottleneck

Results

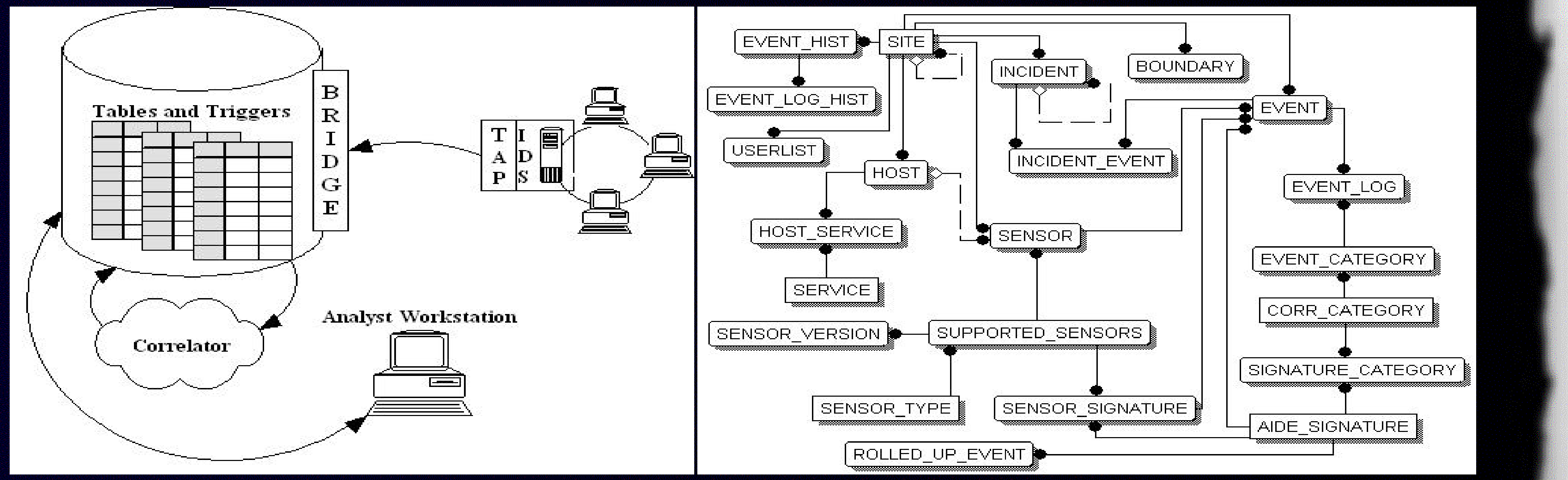
- Demonstrated increased throughput performance with respect to RDBMS
- Demonstrated inconclusive, but seemingly minimal, impact to data retrieval performance
- Demonstrated the effective distribution of the repository to counter the single-point-of-failure problem
- Defined an object-oriented architecture and programming paradigm for future system development

Results immediately generalize to other systems, including the deployed ASIM/CIDDS system currently used operationally by the USAF

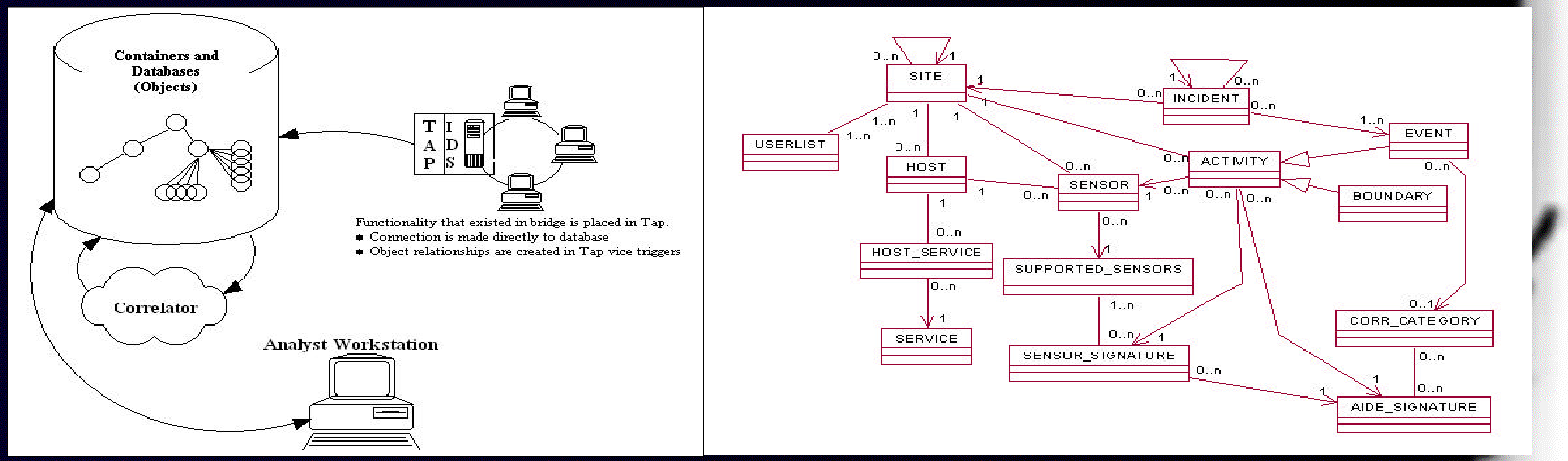
Generic IDS with repository



AFRL's AIDE System Overview

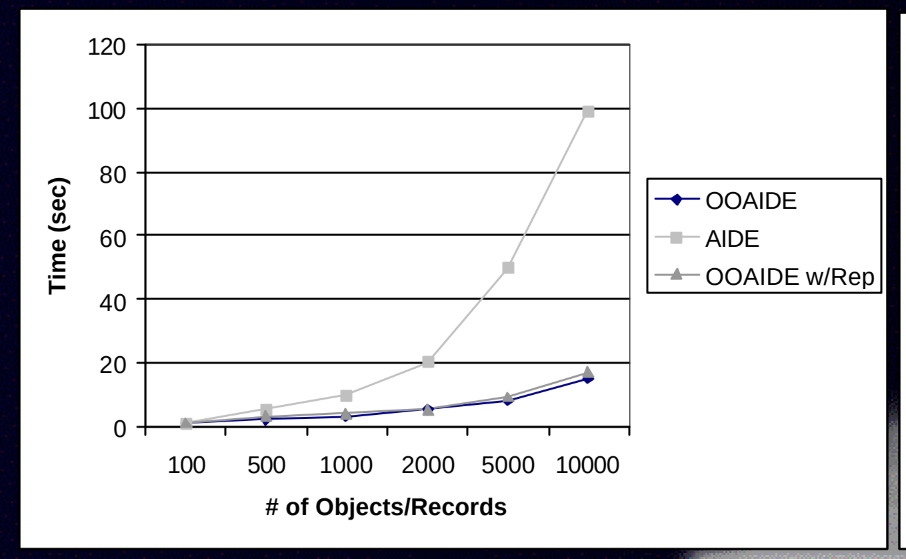


AIDE After Re-Design to DOODBMS

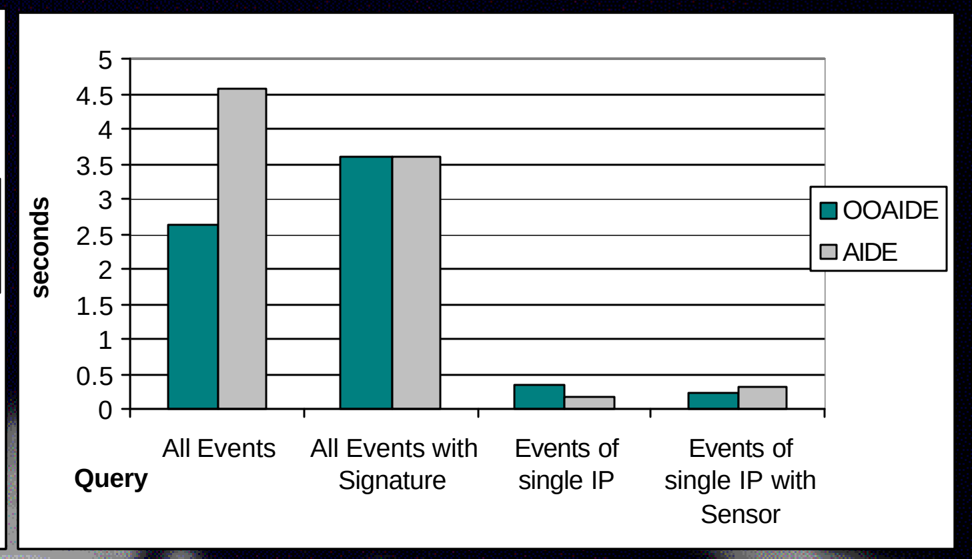


Functionality that existed in bridge is placed in Tap.
- Connection is made directly to database
- Object relationships are created in Tap vice triggers

Insertion Performance Comparison



Query Performance Comparison



Capt Paul D. Williams

Recipient of the Class 01M

Commandant's Award

for Best Thesis

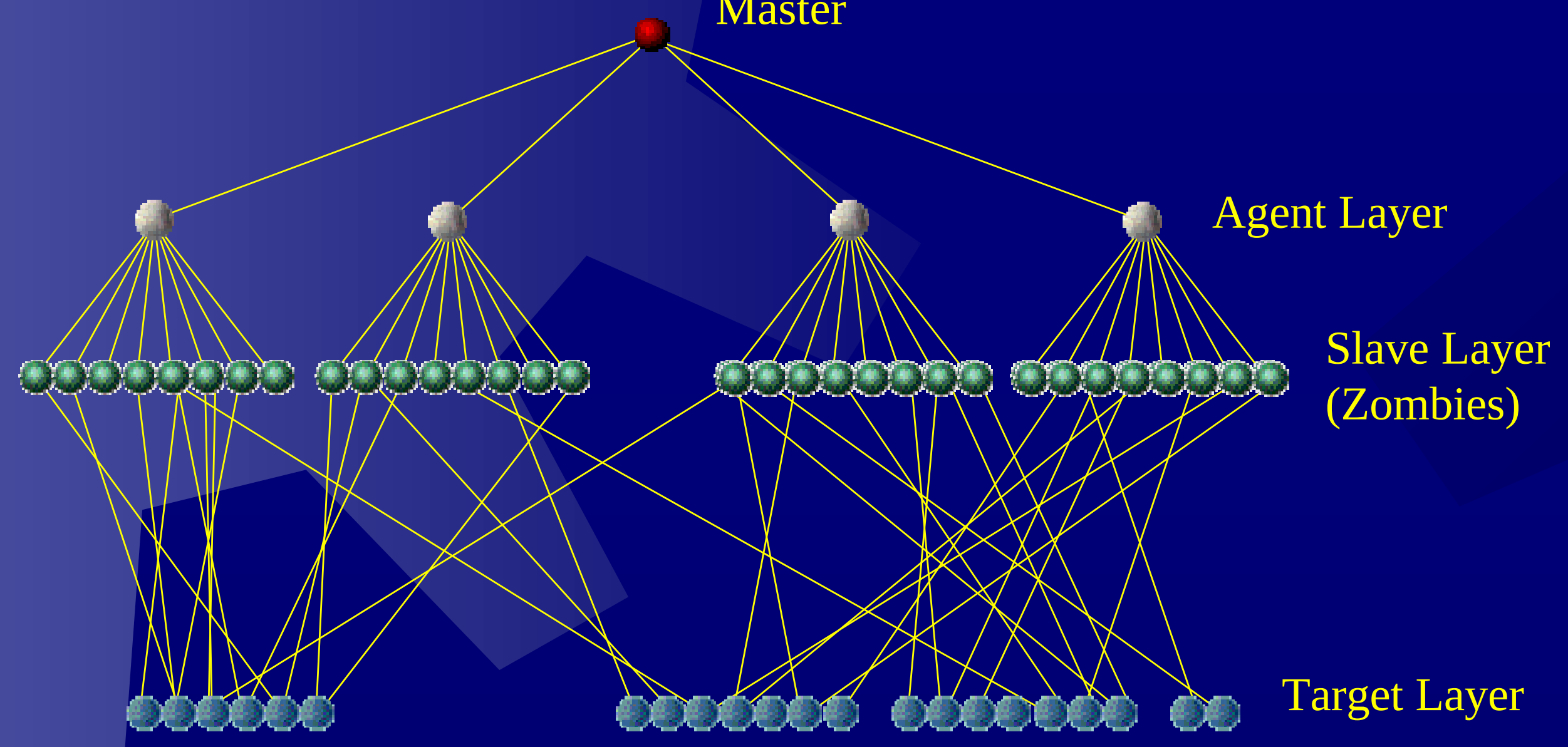
Warthog

Towards a Computer Immune System for Detecting "Low and Slow" Information System Attacks

The Challenge

- **INFOSEC Research Council: Intrusion/Misuse Detection is the top Information Assurance (IA) priority**
- **Office of the Assistant Secretary of Defense (OASD/C3I): Intrusion Detection leads the IA Hard Problems list**
- **National Security Agency: The methodical, structured threat poses the most significant security risk to our National Information Infrastructures**

What is "Low and Slow" ?



- Coordinated, distributed, reconnaissance and penetration attempts perpetrated by a patient, resourceful, structured adversary
- The Master is the adversary's controlling computer
- The Agents and Slaves are computers - usually ***innocent*** - that have been "Trojanized" by an adversary or other agent
  - Current estimate of extant zombies exceeds hundreds of thousands
- Master assigns selected slaves against specific targets - patience is key
  - Individual probes are difficult to distinguish from noise
  - Correlation of probes is nearly impossible
  - Master is well-hidden behind layers of concealment
- Hierarchical structure is also ideal to initiate massive distributed denial of service (DDOS) attacks against Target Layer

Sponsored By

Defensive Information Warfare Branch

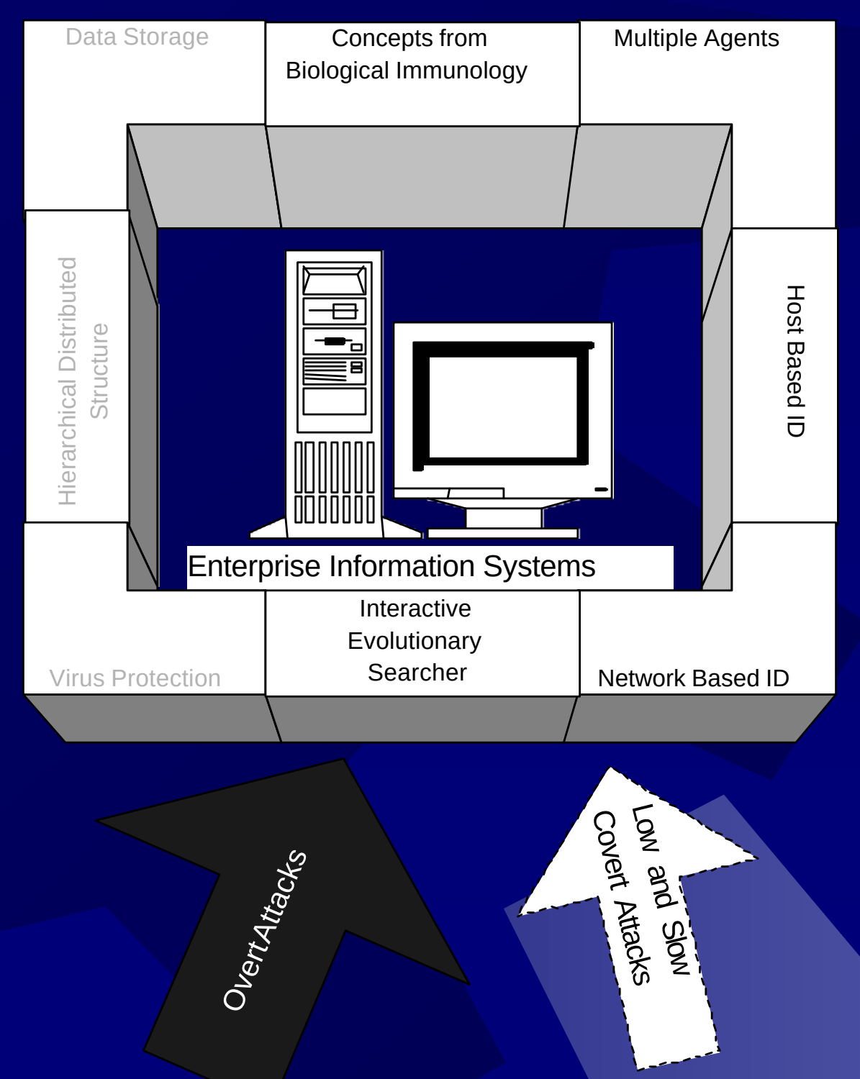
Information Directorate, AFRL/IFGB

Air Force Research Laboratory

The Problem

- Signature-based Intrusion Detection (ID) is reactive
  - Operation depends upon existing signatures
  - Signatures typically created in attack post-mortem
  - Both signature creation and distribution are manual processes
- Signature success depends on generality
  - New attacks are often variations of old ones
  - Problem domain is always changing
  - Problem domain space is enormous
    - 1.94×10^{84} possible events using just 29 packet features

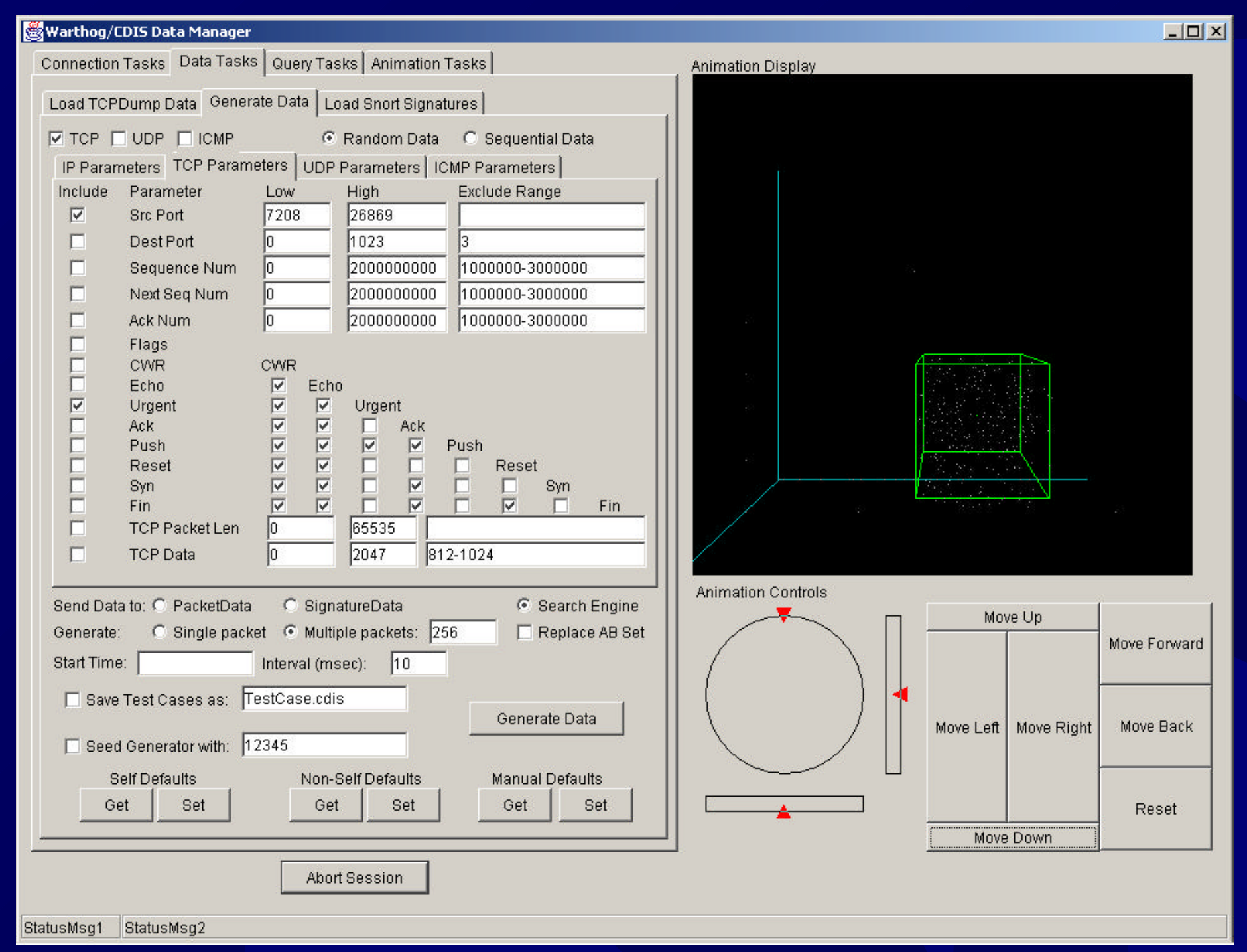
CDIS



The Approach

Computer Defense Immune System (CDIS) ⇒ Warthog

- Build upon AFIT's Computer Virus Immune System
- Integrate several different techniques
  - Computer immunology
    - Develop antibodies through negative selection and maturation
  - Computer virus and intrusion detection
  - AFIT's multi-agent systems engineering (MaSE) methodology
  - Parallel and distributed computation
- Utilize an Interactive Evolutionary Stochastic Search process
  - Genetic Algorithms - coupled with human analyst for search guidance



☑ TCP ☐ UDP ☐ ICMP ● Random Data ○ Sequential Data

Include	Parameter	Low	High	Exclude Range
☑	Src Port	7208	26869	
☐	Dest Port	0	1023	3
☐	Sequence Num	0	2000000000	1000000-3000000
☐	Next Seq Num	0	2000000000	1000000-3000000
☐	Ack Num	0	2000000000	1000000-3000000
☐	Flags			
☐	CWR			
☐	Echo			
☑	Urgent			
☐	Ack			
☐	Push			
☐	Reset			
☐	Syn			
☐	Fin			
☐	TCP Packet Len	0	65535	
☐	TCP Data	0	2047	812-1024

	CWR	Echo	Urgent	Ack	Push	Reset	Syn
Echo	☑						
Urgent	☑	☑					
Ack	☑	☑	☐				
Push	☑	☑	☑	☑			
Reset	☑	☑	☐	☐	☐		
Syn	☑	☑	☐	☑	☐	☐	
Fin	☑	☑	☐	☑	☐	☑	☐

Send Data to: ○ PacketData ○ SignatureData ● Search Engine

Generate: ○ Single packet ● Multiple packets: 256 ☐ Replace AB Set

Start Time: Interval (msec): 10

☐ Save Test Cases as: TestCase.cdis

☐ Seed Generator with: 12345

The Successes

- Provided a formal framework for defining the intrusion detection problem
- Performed simple, single-packet, network-based ID in the context of a CDIS
  - Warthog can separate self from non-self
  - Detects unknown attacks - attacks that were not part of the training data
    - In one test, detected over 98% of 2600+ attacks covering a large number of protocols and techniques
- Defined a search process that couples the skills of human analysts with the raw searching power of an evolutionary algorithm
  - Developed prototype user interface to display and guide search progress
- Provided a means of determining which features are important
  - The collection of successful antibodies will contain a variety of features
  - Induction over those antibodies should extract those features most useful
  - Focusing on those features should improve search and reduce data storage

Preliminary Observations

- Which features are important?
  - IP src, dest addresses
  - IP ID
  - IP TTL
  - TCP Sequence numbers
  - TCP Push, Ack, Syn flags
  - TCP src, dest ports

Biggest Contribution

SOLID Foundation for Continued Research