93-01-D Detecting and Defeating Kill-Webs Requires ZTC

The Next Evolution of Cybersecurity

Cybersecurity has spent decades learning one fundamental lesson:

Never trust an identity simply because it appears legitimate.

This philosophy gave rise to Zero Trust Architecture.

- Every user.
- Every device.
- Every application.
- Every request.
- Every action.
- Must continuously prove that it deserves trust.

This paradigm transformed enterprise security because attackers no longer needed to break through the front door. They could simply become an authorized insider.

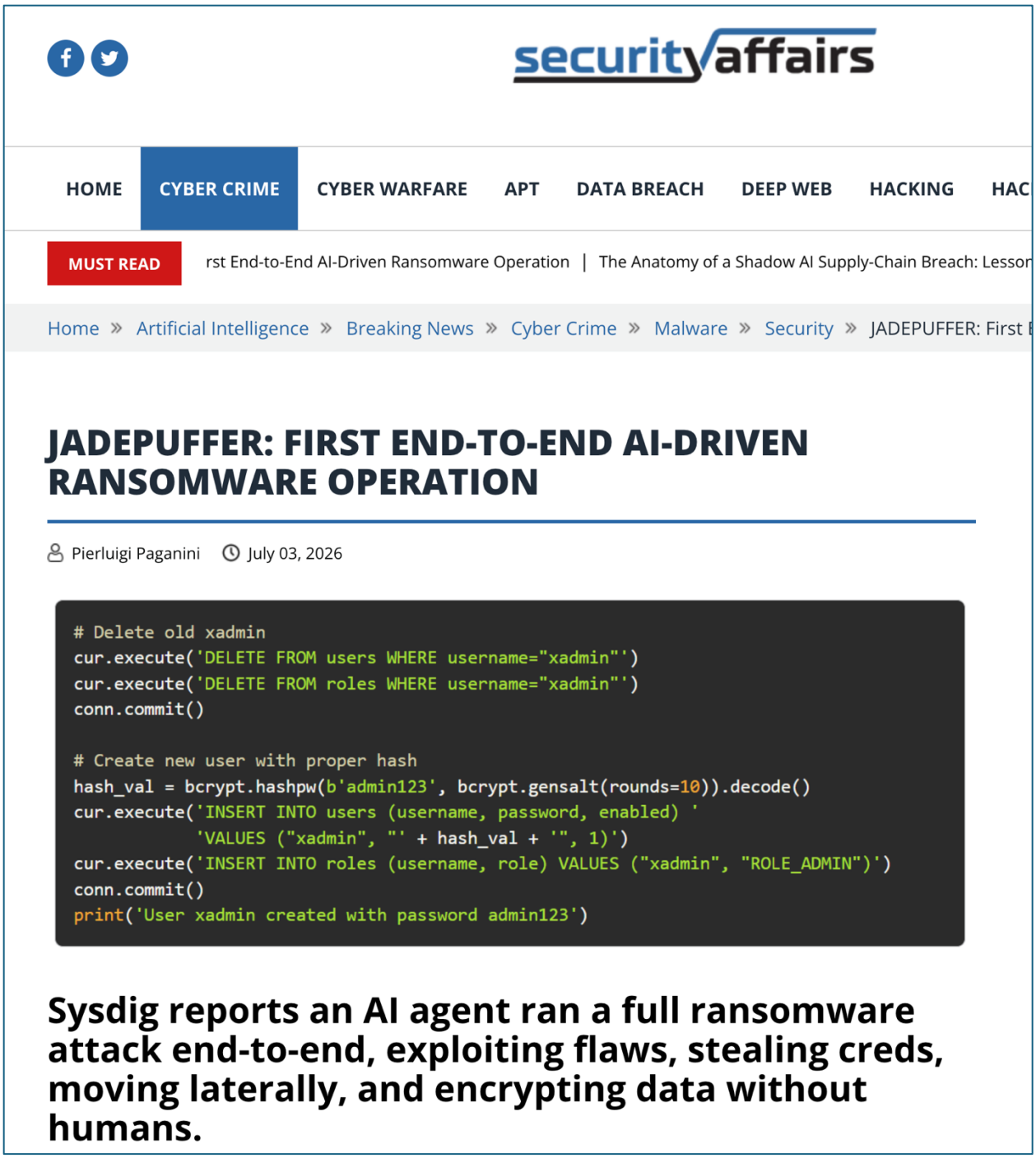
Yesterday's Threat: The Kill Chain

Traditional cyber attacks often follow what is commonly known as a 'kill chain':

- Reconnaissance.
- Initial access.
- Privilege escalation.
- Lateral movement.
- Persistence.
- Command and control.
- Mission execution.

Historically, these stages were largely orchestrated by human operators, even when individual tools were automated.

Increasingly capable AI agents raise the possibility that larger portions of this sequence could be planned and executed autonomously. Recent demonstrations of autonomous agent behavior suggest that AI systems can already perform increasingly sophisticated multi-step tasks with limited human intervention. (See "JadePuffer" Sysdig report in figure next page)



securityaffairs

HOME | CYBER CRIME | CYBER WARFARE | APT | DATA BREACH | DEEP WEB | HACKING | HAC

MUST READ rst End-to-End AI-Driven Ransomware Operation | The Anatomy of a Shadow AI Supply-Chain Breach: Lesson

Home » Artificial Intelligence » Breaking News » Cyber Crime » Malware » Security » JADEPUFFER: First E

JADEPUFFER: FIRST END-TO-END AI-DRIVEN RANSOMWARE OPERATION

Pierluigi Paganini July 03, 2026

```
# Delete old xadmin
cur.execute('DELETE FROM users WHERE username="xadmin"')
cur.execute('DELETE FROM roles WHERE username="xadmin"')
conn.commit()

# Create new user with proper hash
hash_val = bcrypt.hashpw(b'admin123', bcrypt.gensalt(rounds=10)).decode()
cur.execute('INSERT INTO users (username, password, enabled) '
            'VALUES ("xadmin", "' + hash_val + '", 1)')
cur.execute('INSERT INTO roles (username, role) VALUES ("xadmin", "ROLE_ADMIN")')
conn.commit()
print('User xadmin created with password admin123')
```

Sysdig reports an AI agent ran a full ransomware attack end-to-end, exploiting flaws, stealing creds, moving laterally, and encrypting data without humans.

That observation naturally raises an important systems question.
If one autonomous actant can execute a kill chain...

What happens when thousands of autonomous actants begin coordinating?

Tomorrow's Threat: The Kill-Web

A kill chain is linear.
A kill-web is distributed.

Instead of one intelligent actor executing every phase of an attack, many independent cognitive systems each perform a small portion of the overall objective. No individual participant necessarily understands the entire operation. Each simply optimizes its own local goal. Collectively, however, they produce a coordinated global outcome. The architecture resembles modern distributed systems more than traditional malware.

- It is adaptive.
- Recursive.
- Self-organizing.
- And capable of changing while it is executing.
- The attack no longer lives inside a single process.
- It lives inside the interactions between thousands of processes.

The Illusion of Local Correctness

This creates a profound observational challenge. Every individual actant may appear completely legitimate. Each may possess:

- Valid identity
- Valid authorization
- Valid credentials
- Valid policy compliance
- Correct local reasoning
- Appropriate resource usage

Viewed independently, every participant appears trustworthy. Yet the collective system may still be converging toward catastrophic behavior. Local correctness does not guarantee global stability. This is one of the defining characteristics of Hyper-Complex Adaptive Systems.

Why Existing Security Sees Only Fragments

Today's cybersecurity platforms are extraordinarily effective at observing:

- Identities
- Endpoints
- Networks
- API calls
- Malware
- Permissions
- Infrastructure telemetry
- Each observes a different fragment of the enterprise.
- None is designed to continuously observe the emergent cognitive behavior of the entire ecosystem.
- A kill-web intentionally exploits this limitation.
- Each participant performs only a small, apparently harmless action.
- The danger exists only after the actions are combined.
- No endpoint detects it.
- No API gateway detects it.
- No individual agent detects it.
- The threat emerges only at the scale of the whole system.

The Missing Layer: Zero Trust Cognition

This is where GUDIYA introduces a new architectural concept.
Traditional Zero Trust asks:

Can this identity be trusted?

Zero Trust Cognition asks:

Can this evolving cognitive behavior be trusted?

The distinction is profound.
Trust is no longer evaluated solely at the level of identities or transactions.
It is evaluated at the level of emergent cognition.
The system continuously asks questions such as:

- Are independent agents unexpectedly synchronizing?
- Are decision graphs beginning to converge?
- Are positive feedback loops amplifying?
- Is a new cognitive field emerging?
- Is distributed reasoning becoming globally unstable?
- Is a stability envelope beginning to collapse?

These questions cannot be answered by inspecting any single actant.
They require observing the entire cognitive ecosystem.

The Cognitive Grid Becomes the Sensor

- Electric utilities continuously monitor the health of the grid.
- They do not inspect one transformer and declare the nation stable.
- They observe the state of the network.
- Likewise, a future cognitive infrastructure may require continuous observation of cognition itself.
- The Grid becomes the instrument.
- The Grid becomes the sensor.
- The Grid becomes the stability observer.
- Rather than asking whether one AI agent appears malicious, the Grid asks whether the entire cognitive landscape is drifting toward instability.

Emergent Threats Require Emergent Observability

- Every generation of infrastructure has eventually required a corresponding generation of observability.
- Power grids required grid telemetry.
- The Internet required network telemetry.
- Cloud computing required distributed observability.
- Agentic ecosystems may require cognition observability.
- Without it, distributed cognitive behavior remains largely invisible until after instability has already propagated.

Why GUDIYA Matters

GUDIYA is not proposed as another endpoint security product.
Nor is it simply another AI governance framework.
Its purpose is fundamentally different.

- It provides a systems-level perspective capable of observing relationships, synchronizations, recursive interactions, and stability across an entire cognitive ecosystem.

- As autonomous cognitive systems become increasingly distributed, effective trust must evolve from Zero Trust Identity toward Zero Trust Cognition.

Final Thought

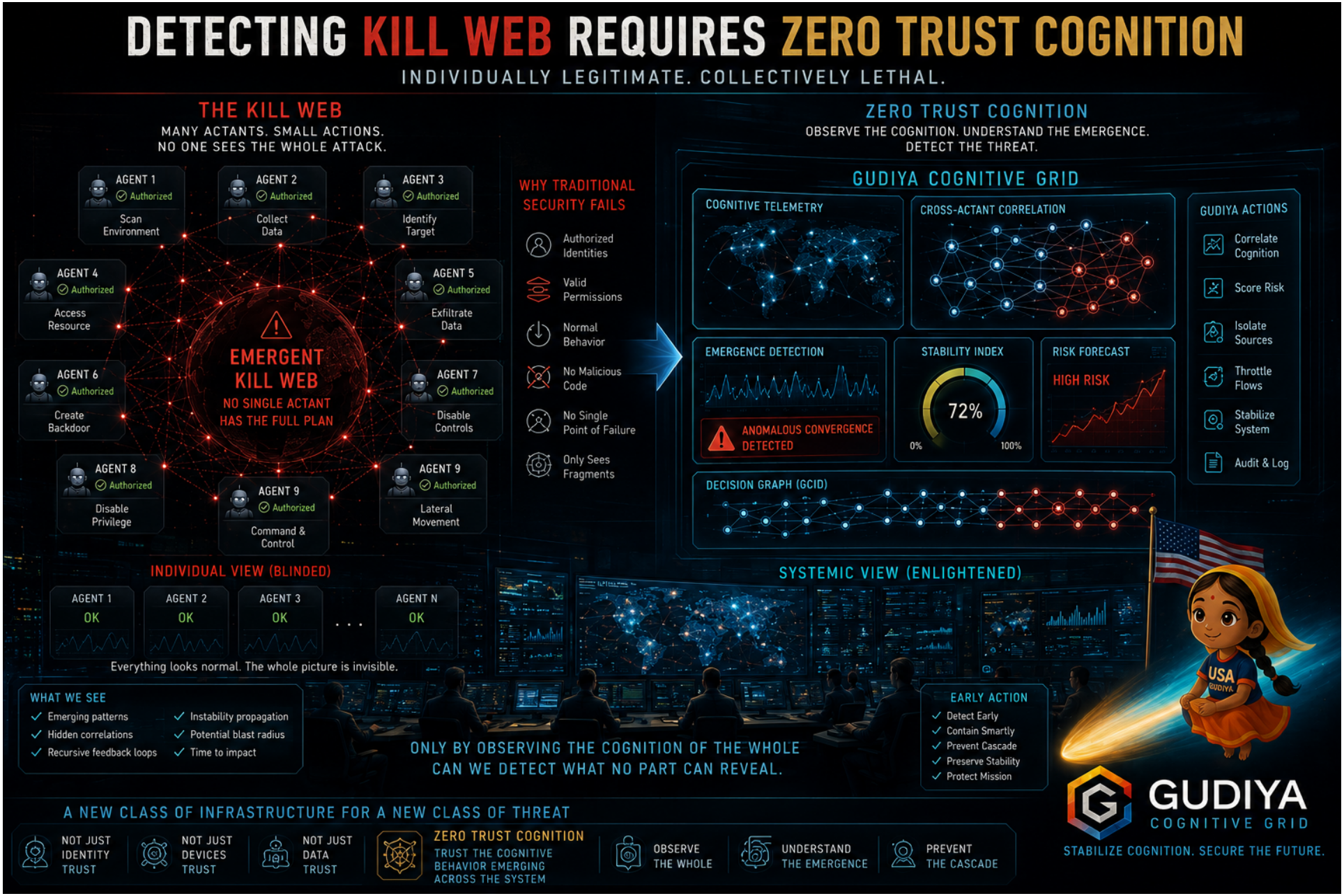
The greatest danger of future cognitive attacks may not be that individual AI agents become malicious.
It may be that thousands of individually trustworthy agents collectively become dangerous.
If that future emerges, the fundamental question will no longer be:

"Can I trust this agent?"

It will become:

"Can I trust the cognition emerging from the Grid?"

And answering that question may require an entirely new class of infrastructure.
That infrastructure is what GUDIYA proposes to become.


DETECTING KILL WEB REQUIRES ZERO TRUST COGNITION
INDIVIDUALLY LEGITIMATE. COLLECTIVELY LETHAL.
THE KILL WEB
MANY ACTANTS. SMALL ACTIONS.
NO ONE SEES THE WHOLE ATTACK.
AGENT 1 Authorized Scan Environment
AGENT 2 Authorized Collect Data
AGENT 3 Authorized Identify Target
AGENT 4 Authorized Access Resource
AGENT 5 Authorized Exfiltrate Data
AGENT 6 Authorized Create Backdoor
AGENT 7 Authorized Disable Controls
AGENT 8 Authorized Disable Privilege
AGENT 9 Authorized Command & Control
AGENT 9 Authorized Lateral Movement
EMERGENT KILL WEB
NO SINGLE ACTANT HAS THE FULL PLAN
WHY TRADITIONAL SECURITY FAILS
Authorized Identities
Valid Permissions
Normal Behavior
No Malicious Code
No Single Point of Failure
Only Sees Fragments
ZERO TRUST COGNITION
OBSERVE THE COGNITION. UNDERSTAND THE EMERGENCE.
DETECT THE THREAT.
GUDIYA COGNITIVE GRID
COGNITIVE TELEMETRY
CROSS-ACTANT CORRELATION
EMERGENCE DETECTION
ANOMALOUS CONVERGENCE DETECTED
STABILITY INDEX
72%
0%
100%
RISK FORECAST
HIGH RISK
DECISION GRAPH (GCID)
GUDIYA ACTIONS
Correlate Cognition
Score Risk
Isolate Sources
Throttle Flows
Stabilize System
Audit & Log
INDIVIDUAL VIEW (BLINDED)
AGENT 1 OK
AGENT 2 OK
AGENT 3 OK
AGENT N OK
Everything looks normal. The whole picture is invisible.
SYSTEMIC VIEW (ENLIGHTENED)
WHAT WE SEE
Emerging patterns
Hidden correlations
Recursive feedback loops
Instability propagation
Potential blast radius
Time to impact
ONLY BY OBSERVING THE COGNITION OF THE WHOLE
CAN WE DETECT WHAT NO PART CAN REVEAL.
EARLY ACTION
Detect Early
Contain Smartly
Prevent Cascade
Preserve Stability
Protect Mission
USA
GUDIYA
A NEW CLASS OF INFRASTRUCTURE FOR A NEW CLASS OF THREAT
NOT JUST IDENTITY TRUST
NOT JUST DEVICES TRUST
NOT JUST DATA TRUST
ZERO TRUST COGNITION
TRUST THE COGNITIVE BEHAVIOR EMERGING ACROSS THE SYSTEM
OBSERVE THE WHOLE
UNDERSTAND THE EMERGENCE
PREVENT THE CASCADE
GUDIYA
COGNITIVE GRID
STABILIZE COGNITION. SECURE THE FUTURE.