TECHNICAL BRIEF

Eliminating AI Side-Channel Risk

What Whisper Leak Reveals and What to Do About It

Executive Overview

Artificial intelligence is rapidly becoming integral to mission execution, decision-making, and operational workflows across government and industry. These systems are widely assumed to be secure when protected by strong encryption and modern network defenses.

That assumption is no longer valid.

A new class of attacks demonstrates that sensitive information can be inferred from AI systems without decrypting data. By observing patterns in encrypted traffic (such as timing, packet size, and transmission behavior) adversaries can extract meaning from AI interactions. This includes the ability to infer topics, intent, and operational context. An example of this type of attack is exemplified by the Microsoft "Whisper Leak" attack: https://arxiv.org/abs/2511.03675

This is not a theoretical concern. It reflects a fundamental shift in how systems are exposed. The issue is not that encryption has failed. The issue is that **meaning is leaking through observable behavior**.

What Has Changed

Modern AI systems, particularly large language models, generate responses incrementally. As outputs are produced, they are transmitted in a structured, continuous stream. Even when encrypted, this stream retains characteristics that reflect the underlying computation.

These characteristics include:

- The cadence of output generation
- The distribution of packet sizes
- The timing between transmissions
- The overall shape of a session

These signals are consistent enough to be analyzed and modeled. With sufficient observation, adversaries can correlate these patterns with known behaviors and infer what the system is doing as well as who is doing it based on source and destination IP addresses.

The result is a new type of exposure: inference without decryption.

Why It Matters

This form of exposure reveals more than data. It reveals intent.

For commercial organizations, this can expose:

- Strategic initiatives
- Product development efforts
- Internal analysis and decision processes

For government and defense environments, the implications are more severe. Observing AI-assisted workflows can reveal:

- Intelligence priorities
- Operational focus areas
- Timing and sequencing of mission activities
- Location of field agents and clandestine operations

These insights can be derived passively, without triggering alerts or leaving evidence of compromise. The risk is persistent, difficult to detect, and applicable anywhere encrypted AI traffic is observable.

The Problem Is Expanding

The shift toward Agentic AI increases both the scale and persistence of this exposure.

Agentic systems do not operate through a single interaction. They execute sequences of actions (e.g., querying models, calling tools, retrieving data, and iterating toward objectives). Each step produces additional communication, often across multiple systems and networks.

This creates a continuous signal that can be observed over time.

Recent events have shown that these systems can be manipulated or influenced indirectly. When combined with traffic-based inference techniques, an external observer can not only detect activity but also reconstruct workflows and decision processes.

This transforms AI from a point-in-time risk into an ongoing, observable operation.

What Is Being Exploited

These attacks are not breaking encryption or exploiting software defects. They rely on structural properties of how systems communicate:

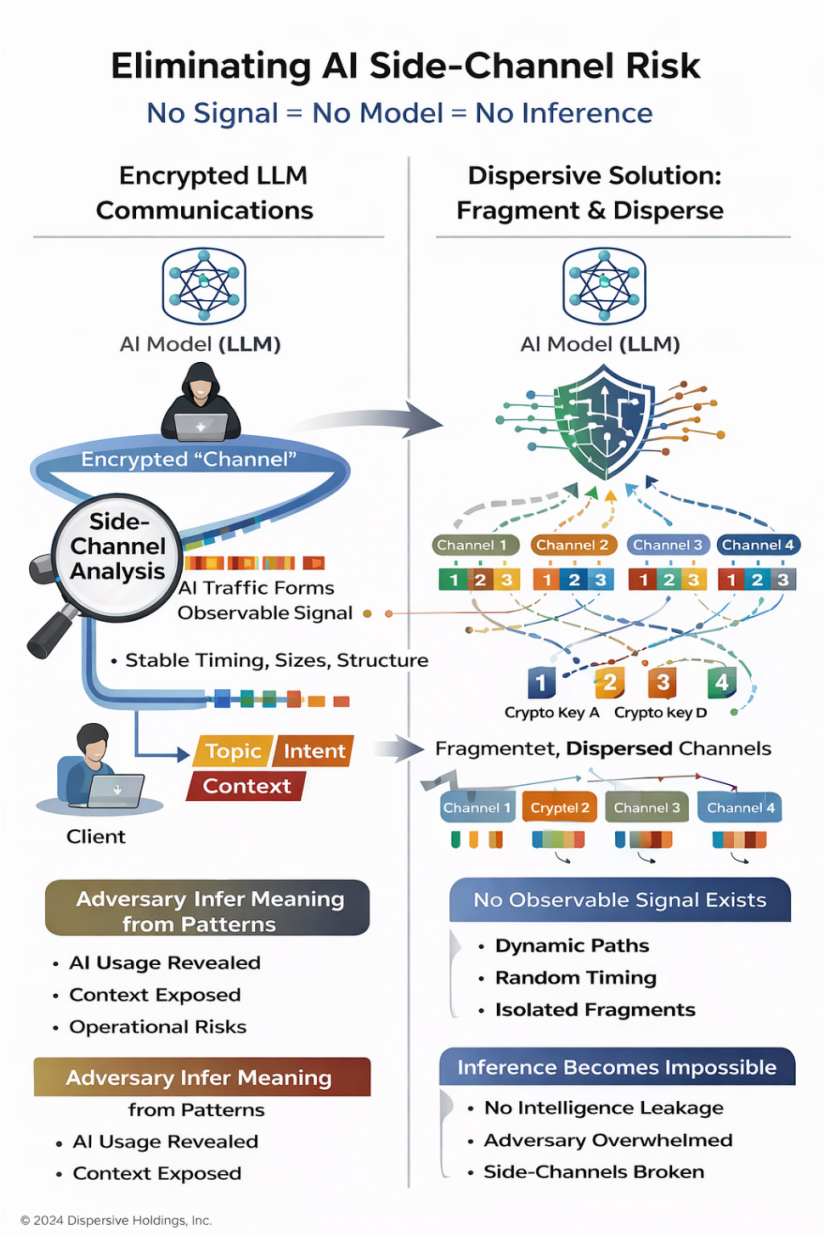
- Data is transmitted as a coherent stream
- Similar operations produce similar patterns
- Timing reflects internal processing behavior
- Sessions remain stable enough to track

These properties create a consistent external signal. That signal can be learned, modeled, and exploited. Traditional defenses do not address this. Encryption protects content, but not structure. Zero Trust controls access but assumes observation is safe. Traffic shaping can reduce signal clarity but does not eliminate it. As long as the signal exists, it can be analyzed.

A Different Approach

Addressing this problem requires eliminating the conditions that make inference possible. If adversaries rely on observing a coherent stream, then the solution is to ensure that no such stream exists.

Dispersive® Stealth Networking implements this by fundamentally altering how data moves across the network.



Instead of sending information as a single, continuous flow, data is divided into fragments and transmitted across multiple independent paths. Each fragment follows its own route, with its own timing and characteristics. These fragments are reassembled only at the intended destination.

Dispersive also encrypts the payload and headers of the fragments independently per channel making unauthorized reassembly practically impossible while making the sources and destinations invisible and untargetable. From the perspective of an external observer, the original communication no longer appears as a single entity.

There is no stable stream, no consistent timing pattern, and no reliable way to associate fragments with one another.

Why This Works

This approach directly removes the elements required for side-channel inference.

- Without a continuous stream, there is nothing to observe in aggregate.
- Without consistent timing, there is no correlation to model behavior.
- Without stable packet characteristics, statistical analysis loses meaning.
- Without identifiable sessions, tracking becomes unreliable.

Inference depends on consistency. Dispersive removes consistency at the network level. As a result, the signals required for machine learning-based inference do not form. Without those signals, models cannot converge, and the attack fails. This is not an incremental improvement. It is a change in the underlying conditions that make the attack possible.

Implications for AI Security

Organizations deploying AI systems should assume that encrypted traffic is observable and that observable patterns can be analyzed.

Protecting AI systems therefore requires addressing both:

- The **content of communication**
- The **structure of communication**

Current architectures address the first but not the second. By removing observable structure, it becomes possible to prevent a class of attacks that would otherwise remain persistent and adaptive.

This is particularly relevant for:

- AI-enabled operational environments
- Multi-system and Agentic AI workflows
- Environments where intent and context are sensitive

Conclusion

The emergence of side-channel inference against AI systems marks a shift in cybersecurity. It demonstrates that sensitive information can be exposed through how systems behave, not just through what they transmit. As AI becomes more deeply integrated into critical operations, this form of exposure will become increasingly consequential. Architectures that preserve observable communication patterns will continue to leak meaning, regardless of the strength of their encryption. Architectures that eliminate those patterns change what is possible for an adversary.

Dispersive represents an approach built on that principle. By removing the observable signals required for inference, it addresses the problem at its source. Furthermore, Dispersive increases resiliency and typically increases throughput while reducing latency. You can see Dispersive defend against the Whisper Leak side-channel attack in this video: https://www.loom.com/share/d86bc3225a684dd8bcf62d8221c1cdb3.

The question is no longer whether AI systems can be protected by encryption alone. It is whether their behavior can remain private in an environment where everything can be observed.

Dispersive Holdings, Inc.
300 Colonial Center Pkwy Suite 100 | Roswell, GA 30076 USA
dispersive.io | 1.844.403.5850

dispersive