Episode Book

EPISODE #6115

Claude Cracked the NSA's 1997 Curve Seeds

PUBLISHED	19 October 2026 at 10:55 UTC
DURATION	21:10
LISTEN	myweirdprompts.com/episode/claude-nsa-curve-seeds-1997-45060/
HOSTS	Herman, Corn, Larry, Hilbert
SCRIPT MODEL	anthropic/claude-opus-5.5
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.10
BOOK GENERATED	19 October 2026 at 17:43 UTC

SYNOPSIS

For 25 years, nobody could explain where the NSA's elliptic curve seeds came from — until Anthropic says Claude traced them to a Unix clock in 1997. We walk through the actual method: a missing top bit, an old L rand 48 generator, and lattice reduction straight out of the late-eighties textbook. Then the harder question: if an unreleased internal model did this, what does that mean for Mythos sitting behind a velvet rope and Fable out in the wild?

TRANSCRIPT



Herman

Twenty-five years of people muttering the word backdoor, and the answer turns out to be somebody's computer clock in the summer of nineteen ninety-seven. I am delighted.



Corn

Daniel caught the same story. Anthropic says Claude worked out where the NSA got the seed numbers behind the standard elliptic curves, the ones half the internet signs things with, and it did it in under half an hour. He wants it checked carefully, because it is the kind of headline that grows in the retelling. What actually got solved, how the model did it, and what it says about where these systems are now. Then the bit he is really chewing on. Anthropic kept Mythos behind a velvet rope as too dangerous to release, and yet something that just cracked an NSA cryptography puzzle Anthropic says an unreleased internal research model did it, no name, no version number. So here's the tension. Mythos sits behind a velvet rope as too dangerous to release, Fable is out there as a Mythos-class model anyone can pay for, and now an internal model is being shown off for cracking an NSA cryptography puzzle. Where exactly is the line, and does this sit on the safe side of it?



Herman

We'll get to the line, because I think this story is a good test of it. The puzzle first, though. It is a lovely puzzle.



Corn

Go on. I can see you've brought the whole binder.



Herman

Half the binder. In two thousand, NIST published the curves in the federal signature standard. P two fifty-six is the famous one, and P three eighty-four next to it. Each curve's constants are supposed to be verifiably random. You take a seed, run it through SHA one, and the curve falls out. The point is that nobody can have hand-picked a curve with a secret weakness, because the hash chose it.



Corn

Unless you hand-pick the seed. Try a few million seeds, keep the one you like.



Herman

That's the whole suspicion, Corn. The seeds came from the NSA with no explanation. Just long strings of hex. After the Dual EC business, where an NSA-backed random number generator really did look like it had a trapdoor, people were not in a trusting mood about NSA numbers.



Corn

Fair enough. The NSA used up its benefit of the doubt on that one.



Herman

There was an explanation, sort of. The seeds were generated by an NSA mathematician called Jerry Solinas back in ninety-seven. He reportedly said he had hashed English phrases, something along the lines of Jerry deserves a raise, and then forgot the exact wording. People tried. Nothing matched. Solinas died in early twenty twenty-three, and that autumn the cryptographer Filippo Valsorda put up a bounty for anyone who could find the phrases. Twelve thousand two hundred and eighty-eight dollars, tripled if the winner gave it to charity.



Corn

Twelve thousand two hundred and eighty-eight. Only a cryptographer prices a bounty in powers of two.



Herman

It works out to twelve kibidollars, yes. And it sat there. Password-cracking people threw hardware at it, because SHA one is fast to brute-force. Nothing.



Corn

Because they were all looking for a sentence.



Herman

Which is the first clever thing in this story, and it wasn't the model's. The Anthropic employee who asked deliberately left out the clue. No mention of the phrase, no Jerry deserves a raise. Just, here are the seeds, how were they made?



Corn

Keep the witness away from the misleading testimony.



Herman

And the model's first move was a guess. It reckoned that software in the late nineties probably used one of the weak random number generators that shipped with C libraries back then. Not a hash of anything. Just whatever the computer's built-in random function spat out.



Corn

That's the part I want to understand, actually. Everyone else read the seeds as the output of a hash, and a good hash output looks like noise. If it came from a cheap generator instead, wouldn't that look like noise too?



Herman

Mostly, yes. Mostly. So it wrote a scanner. It slid along each seed one bit at a time and checked whether neighbouring chunks could be consecutive outputs of one of those old generators. It tried thirty-four combinations of generator and slicing. And it spotted something small. In four of the ten prime-curve seeds, the top bit of every chunk was zero.



Corn

Which a hash would never bother doing.



Herman

Which a hash does by coincidence about one time in five thousand eight hundred and ten, by Anthropic's arithmetic. But there's a classic Unix function called L rand forty-eight that only ever hands out thirty-one bits. Pack thirty-one bits into a thirty-two-bit slot and the top bit is always zero.



Corn

So the fingerprint is a missing bit. The thing that gave it away is something that isn't there.



Herman

Then the heavy lifting. L rand forty-eight is a linear congruential generator. It keeps forty-eight bits of hidden state, multiplies by a fixed constant, adds a small number, and shows you only the top slice. For the prime curves, whoever wrote the program kept just a sliver of each output, about eight bits. Recovering the hidden state from slivers like that is a known problem, and there's a known tool for it, lattice reduction, the LLL algorithm.



Corn

That sounds like the part a model would have had to invent.



Herman

It didn't have to. The maths for breaking truncated generators like this goes back to the late eighties. It's textbook. And that matters for Daniel's question about intelligence, because the clever part wasn't inventing a new attack. It was suspecting that this was the kind of problem the old attack fits.



Corn

Then once you have the state...



Herman

You run the generator backwards. Every L rand forty-eight run starts from a state set by its seeding function, and a lot of programs in the nineties seeded it with the current time. So you walk back to the start, read off the seed, and it's a Unix timestamp. Every prime-curve seed lands somewhere between May nineteen ninety-seven and April nineteen ninety-eight.



Corn

I'm guessing P two fifty-six, the one in my browser right now, got made carefully. Office hours, a second pair of eyes.



Herman

Two in the morning, Corn. The run started on the sixteenth of August, nineteen ninety-seven, a little before two in the morning, UTC. The published seed is the one thousand eight hundred and forty-second candidate that run produced.



Corn

Hm. The most important curve on the internet was made by someone who should have gone to bed.

L

Larry

This is a public service announcement. Every night, while your family sleeps, your clocks are ticking. Each tick is a timestamp, and each timestamp can be traced. Somewhere, right now, a stranger with a lattice could be working out exactly what time your toaster was born. Do not wait for that knock on the door. Protect your household with Tick Fog, the only blanket woven from pure decorrelated wool. Drape it over any clock and the time becomes genuinely unknowable, even to you. Missed appointments are not a side effect. They are proof it is working. Tick Fog. Because the hour you were seeded is nobody's business.



Herman

...which also explains the binary curves, or nearly. Their seeds decode to timestamps in February nineteen eighty-four, which makes no sense, until you double them and get the seventeenth of April, nineteen ninety-eight. So that program probably halved the clock somewhere.



Corn

A bug, in other words. A bug from twenty-eight years ago, found by reading its fingerprints on the output.



Herman

And it picked up the base points of a few curves too, the starting points you do arithmetic from. P one ninety-two and five related curves from the ANSI standard came from another old generator, seeded with small numbers starting at twenty thousand.



Corn

Herman, I'm going to do the bit you hate. Under half an hour. Says who?



Herman

Says Anthropic. Nobody independently timed it. But here's what makes me relaxed about it, Corn. They published the code, public domain, and it regenerates the seeds. You don't have to trust the half hour to trust the result. Run it, and either P two fifty-six comes out or it doesn't.



Corn

That's the most falsifiable press release I've ever seen.



Herman

It's a ten-person team on the byline, including Nicholas Carlini, who has spent years attacking machine learning systems for a living. These are not people who need a puff piece.



Corn

I don't believe they found everything, though. Nobody finds everything in half an hour.



Herman

They didn't, and they list what's missing. One binary curve seed, B one sixty-three, is still unexplained. The base points for P two twenty-four, P two fifty-six, three eighty-four and five twenty-one are still unexplained. And there's a strange one. On the P two fifty-six run, candidate eight hundred and seventy-six would also have made a perfectly usable curve, and it was skipped. Anthropic says plainly that it doesn't know what rule was applied.



Corn

So the conspiracy people still have candidate eight hundred and seventy-six.



Herman

They do. It's not much of a hideout. If you were rigging the curve, you'd be cherry-picking from millions of seeds, not stepping over one candidate in a run off the system clock. And the post is blunt about this. Not a break, not a new attack, nothing anyone needs to change. If anything the curves look more trustworthy now. What about you, Corn? Does this make you trust P two fifty-six more, or just trust the NSA's sysadmin less?



Corn

More. I trust the curve more. Nobody was being sinister, somebody was running a program late at night with the default random function. From our own side, I'll take sloppy over sinister any day.



Herman

There's a lovely irony in it, too. The whole point of verifiably random was a receipt. Here's the seed, here's the hash, so you know nobody picked the curve. Except the receipt only works if you can explain where the seed came from, and for twenty-five years nobody could. It was a receipt with the shop's name torn off.



Corn

And now it's been taped back on, by a machine, a quarter of a century late.



Herman

Retroactively, yes. What I can't tell you is why anyone was running it at two in the morning. My guess is a batch job left going overnight, but that's a guess. Nothing in the post says.



Corn

Or a mathematician who'd just been told he wasn't getting a raise.



Herman

Ha. Don't.



Herman

And the bounty, all thirty-six thousand eight hundred and sixty-four dollars after the charity tripling, goes to the nonprofit behind Let's Encrypt. Which is a lovely place for NIST curve money to end up.



Corn

The release question, then. The model that did this is internal. Is anything stopping the public Mythos-class one from doing the same?



Herman

Let me get the timeline straight first. Mythos was disclosed in April and held back because it was finding and exploiting bugs in every major operating system and browser. Project Glasswing gave it to a few dozen big defenders, and by May they'd found over ten thousand high or critical vulnerabilities. Then in June Anthropic shipped Fable, which the reporting describes as the same underlying model with safeguards on top, and kept Mythos itself restricted.



Corn

We covered the bit where the government switched Fable off three days in, and frankly that episode was about who gets to decide what a model is allowed to do. This is the other side of it. What the safeguards actually let through.



Herman

The safeguard is the interesting part. When Fable's classifiers flag a request as cybersecurity, or bio, or chemistry, it doesn't refuse. It quietly hands the request to a less capable Opus model. So the Mythos-class brain is public, but its hands get swapped out whenever you ask for the dangerous thing.



Corn

What do you think, then? Would this question get swapped out?



Herman

I genuinely don't know. Recover the internal state of a random number generator from its output... a classifier could easily read that as attack work, because it is attack work. It's exactly how people break badly seeded session tokens and old gambling machines. My guess is a public model would sometimes get bumped to the Opus fallback on this, and nobody outside Anthropic knows whether the fallback would get there.



Corn

Here's where I come down on the line, though. What got held back was never just being clever. It was a particular kind of clever. Point it at live software and it hands you a working break-in. This was archaeology. The output is a date in nineteen ninety-seven and a sentence saying the curves are fine. Nobody's bank account is less safe tonight.



Herman

That's the distinction Anthropic is drawing, and I mostly buy it. Withholding Mythos was about scale and targets. A model that finds ten thousand holes in software running in hospitals right now is a different object from a model that explains a twenty-five-year-old seed.



Corn

Mostly?



Herman

Mostly, because the skill underneath is the same skill. Spot that something which looks random isn't, guess the generator, rebuild its hidden state. The NIST seeds are harmless because nineteen ninety-seven is over. Point the same move at a product that shipped last month with a lazy random number generator and it isn't archaeology any more. The tension is real. Whether it's dangerous depends almost entirely on what you aim it at, and the model itself can't always tell.



Corn

Which is why they had to build the swap-out instead of a smarter refusal.



Herman

Here's what I'd ask you, though. If you were Anthropic, would you have published this at all?



Corn

Yes, and loudly. It's a capability demo where the only casualty is a conspiracy theory. Here's what our internal models can do, here's a result anyone can check, and here's the harmless end of it.



Herman

That's how I read it too. It's a demonstration as much as a finding.



Corn

Then what does it tell us about the intelligence? You've been itching for this part since the top.



Herman

I have. Three things. The leap was the hypothesis. Thousands of very good people spent twenty-five years assuming the seeds were hashes, because the man who made them said so. The model, kept away from that story, made a different guess that fit the evidence. The second is that it knew which tool fits which puzzle, a scanner first, then lattice reduction, then running the clock backwards, without anyone laying out the route. The third is speed. Whatever you think of the half hour, the bounty sat unclaimed for three years.



Corn

Let me be the slow one, as usual. A human expert told to ignore the Solinas story and suspect a cheap generator could have done this too.



Herman

They could. Would you have guessed a cheap generator, Corn, if someone had handed you the hex and kept quiet about Jerry?



Corn

No. I'd have trusted the dead man's story like everybody else. That's rather the point.



Herman

Every individual step is known. Nobody aimed the expertise at it, partly because the official story pointed everyone the other way. What the model contributes is that it's cheap to aim. You can point it at a thousand dusty mysteries and see which ones crack.



Hilbert

Random. You both keep saying random. The permits office in my area says the residents' recheck is random, and I've been drawn eleven months running.



Herman

Eleven in a row would be a very unlucky coin.



Hilbert

Tortoise. Mrs. Pelleg at the residents' permits office keeps a tortoise called Bernard in a shallow tray with a grid painted on the bottom. Every resident on the street gets a square. On the first Tuesday of the month she puts Bernard in the middle, and whichever square he stops on has to come in and prove they still exist. He stops on my square. Every month. I go in on the Wednesday with my birth certificate, my lease, and a signed letter from a man called Avner who keeps pigeons on the roof opposite and can see me hanging up washing.



Corn

The pigeon man is the witness.



Hilbert

The pigeon man is the only witness they'll accept. The man downstairs offered to sign one and they said he can only vouch for my ceiling. Avner charges me forty shekels a letter and he puts the date on it in pencil so he can rub it out and reuse it if the office loses it, which it has done twice. Two of Avner's pigeons have residents' permits of their own. Neither has ever been drawn. This month Mrs. Pelleg has started making me sign for Bernard's lettuce as well, because the lettuce comes out of a residents' fund and I'm the resident who's always there. The pen at that counter is on a chain made for somebody with fingers, so it takes me a while. She waits. She always waits. She turns the form round so I don't have to reach.



Herman

The tray is the generator, then. The grid's fixed, so whatever Bernard does is the only randomness in the system.



Hilbert

Bernard is not a generator, Herman. He's a tortoise with a pension. He's been at the office longer than the photocopier, and he has a desk plaque. In September he was away on a course in Haifa, and the draw was done by his understudy, a smaller tortoise the office calls Also Bernard. Also Bernard stopped on my square too. What I want to know is this. If



Hilbert (continued)

your model can look at a list of numbers and say, these came from a clock at two in the morning on the sixteenth of August, can it look at eleven months of Bernard and say where he's going to stop in November? I've written down every square since last December. I've also written down where all the other residents' squares are. I'm not saying why that's relevant.



Corn

You don't have to.



Hilbert

I'm asking whether it's possible. In principle. For a tortoise.



Corn

Hilbert, you've been there every month to sign for the lettuce, have you? So when Mrs. Pelleg puts the lettuce down before the draw, where does it go?



Hilbert

On my square. That's where I'm standing when I sign. It's the corner nearest the counter.



Herman

Then you don't need a model. The hidden state in that tray is the lettuce, and the lettuce is wherever you are.



Corn

You're the system clock, Hilbert. You've been seeding him.



Hilbert

Mm. I'll stand at the other end of the counter in November. Avner will still want his forty shekels.



Corn

Everybody spent twenty-five years interrogating the clever machinery, the hash, the maths, the NSA, and the weak spot was a plain clock sitting next to it the whole time. I do wonder how many other old puzzles are like that, now that checking one costs half an hour.



Herman

Some of those answers will be lovely, like this one. Some of them will be live software, and that's the part Anthropic is still keeping behind the rope.



Corn

Thanks to our producer, Hilbert Flumingtop, for keeping the show on the air.



Herman

This has been My Weird Prompts. If you enjoyed this one, leave us a review on Spotify, it genuinely helps people find the show.



Corn

We'll be back soon.