輝達H20：讀懂何爲晶片後門及潛在風險

美國宣布解除對AI晶片龍頭輝達（Nvidia）H20晶片（芯片）銷往中國的禁令後不久，中國國家網信辦於7月31日突發聲明，宣布已約談輝達，要求其針對擬銷往中國的H20算力晶片存在漏洞與後門安全風險的問題，作出說明並提交相關證明材料。

此事不僅令人關注輝達在中國市場面臨的新挑戰，更將焦點轉向晶片如何被植入後門系統，以及網信辦所指控的「追蹤定位」或「遠端關閉」等功能的可能性。

輝達此後在官網發文稱，輝達晶片中沒有後門、終止開關和監控軟體，並稱「這些絕不是構建可信系統的方式，也永遠不會是。」

晶片的「後門系統」（backdoor system），泛指一種隱藏的機制，允許未經授權的訪問繞過標準安全措施。

專家向BBC指出，在某些極端情況下，這類後門構成嚴重的安全威脅，因其可能讓攻擊者取得控制權、竊取數據或操控設備功能，甚至實現遠端監控。此類後門可能使攻擊者在不被察覺的情況下，監視用戶行為、竊取敏感資料（如AI模型參數或用戶數據），甚或干擾關鍵基礎設施的運作。例如，在軍事或金融應用場景中，後門可能引發災難性後果，進而引發對晶片供應鏈安全的廣泛質疑。

「後門」定義存在爭議

科技安全專家向BBC表示，關於「後門」的定義，確實存在爭議與諸多隱憂。業界對後門的認知分為技術性和惡意兩種觀點：部分被標記為後門的功能，可能是設計中的正常部分，卻也可能因疏忽或惡意被利用。

美國網路安全專家、知名科技網站「安全硬體」(Security-Hardware.com)總監喬?費茲派崔克（Joe FitzPatrick）向BBC中文分析指出，每家製造商的每款晶片皆具備調試功能，儘管實現方式與細節各異，但當客戶需探究設備故障原因時，製造商希望能提供協助。他進一步闡釋，這些調試功能通常用於「診斷硬體問題」或優化性能，例如透過JTAG（聯合測試行動小組）介面存取晶片內部狀態。他強調，部分研究人員曾誤將這些調試功能視為惡意後門，但後來證實其為未記錄的調試功能，而非蓄意設計的惡意機制。

儘管如此，他也坦言，這些調試功能可能存在漏洞，或被製造商濫用，但僅因這些必要功能的存在，便推斷其具「惡意」意圖，並不公允。

費茲派崔克進一步解釋，區分設計缺陷與故意植入的後門，需仰賴精密的技術分析，而非僅憑臆測。他強調，另一類威脅為「硬體木馬」，學術研究對此頗多探討，聚焦於單一行為者可能造成的破壞。例如，一名設計者能否在不被其他設計者、製造廠或測試流程察覺的情況下植入後門？製造廠能否修改設計，繞過驗證流程？釐清這些問題需耗費巨大心力。

費茲派崔克稱，H20晶片專為中國市場設計，符合美國對中國市場的出口管制要求。但這塊晶片複雜的供應鏈（涉及台積電等代工廠和多家第三方IP供應商）增加了後門風險。他向記者表示，說到底，輝達的設計流程高度依賴全球合作，這使得確保每一環節的安全性成為挑戰，尤其在當前地緣政治緊張的背景下。但他也補充強調稱，昂貴的高階AI晶片從設計到生產都要經過不同公司及部門的層層把關，要加入後門並不容易。

根據費茲派崔克的研究，已有諸多理論與原型案例，例如在外部刺激觸發前禁用附加功能、在非無線設備中加入有限無線功能，或透過功耗、性能變化等外部可見特徵洩漏資訊。此外，備受關注的「殺手開關」（kill switch，泛指緊急停止裝置，用於在意外情況下迅速關閉或中斷機器、設備或程序）亦是焦點之一。

費茲派崔克向BBC強調，他未曾聽聞任何機制能透過無線訊號觸發獨立設備自毀。然而，他也指出，美國與中國均有諸多設備需「啟動」才能完整運作的案例，多數在系統層級而非晶片層級，例如遊戲主機、智慧手機或智慧無人機。在晶片層級，某些晶片可能內建數位版權管理（DRM）機制，若未獲得授權，晶片可能無法正常運行，此機制因而被部分批評者視為潛在的「殺手開關」。

以此次風波主角輝達為例，作為一家無晶圓廠的積體電路公司，輝達設計包含第三方矽晶的系統單晶片（SoC），並外包製造。費茲派崔克向BBC表示，輝達的客戶基於這些晶片打造系統，因此每一相關方——設計者、製造商、客戶——均有機會篡改設計，但同時也需密切合作以確保產品正常運作。

他向BBC指出，由於晶片供應鏈的複雜性，注入惡意後門極為困難。此舉需由單一方完成，且必須繞過所有其他方的檢測。

如何植入晶片後門？

晶片後門可能在一片晶片生命週期的不同階段被引入——設計、製造或後期生產。以下是可能的途徑，揭示了晶片供應鏈的脆弱性：

一、設計階段：後門可由原始設計者或工程師嵌入晶片設計中。例如，可在晶片架構中添加隱藏電路或邏輯，允許特定指令觸發特權訪問。此舉可能受到政府或商業競爭對手的驅使。例如，設計者可能在晶片的寄存器傳輸級（RTL）程式碼中插入隱藏邏輯，於特定輸入序列下啟動未記錄的功能，從而允許遠端控制或數據洩漏。

此外，駭客可能危害用於設計晶片的軟體工具，例如透過在設計工具鏈中插入惡意程式碼，在設計團隊不知情的情況下，將後門植入晶片藍圖中。現代晶片常使用來自第三方的知識產權（IP）模組，若這些模組包含惡意邏輯，亦可能引入後門。特別是晶片設計公司，通常將其設計授權予多家製造商，若提供「受損的設計」，後門可能廣泛傳播。

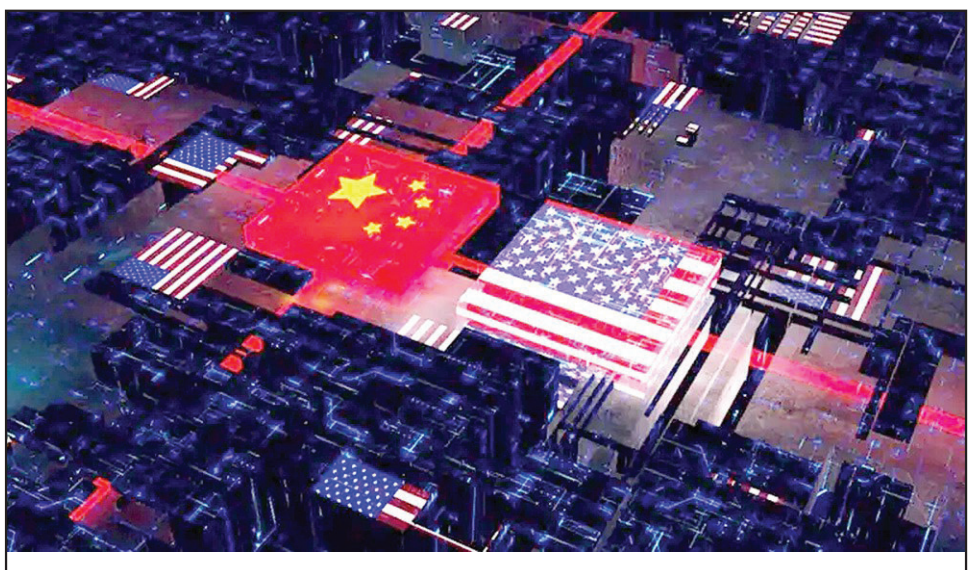
二、製造階段：大多數晶片由第三方代工廠（通常位於海外）製造。代工廠員工可能透過更改晶片佈局（例如修改光罩）來植入後門，例如在特定條件下啟動的電路。

此外，硬體木馬，即對晶片的惡意修改（例如添加微小電路），亦可能成為後門。密西根大學過去的研究顯示，硬體木馬可設計得極為微小且隱秘，幾乎無法透過X射線或光學檢查檢測，增添防範難度。

在製造過程中，後門可能利用模擬屬性或側通道（例如功率變化）變得隱秘，難以透過傳統功能測試發現。

三、後期生產：晶片通常依賴韌體（firmware）運行。韌體是指用來控制底層晶片裡硬體的軟體，如相機或HDMI，並且通常儲存在特定硬體，通常由系統廠提供。惡意更新韌體可能引入後門，允許遠端訪問或控制。此外，後門可能在製造後透過物理修改硬體（例如安裝惡意晶片）被植入。例如，愛德華·斯諾登（Edward Snowden）於2013年聲稱，根據外洩的美國國家安全局文件，該機構曾在運輸過程中攔截硬體，於伺服器與路由器中植入後門。

最後，晶片製造商有時會在晶片中保留預設帳戶、未記錄的遠端訪問系統或除錯模式以進行測試。若未移除，這些可能成為後門。專家強調，商業晶片中硬體後門的公開驗證案例極少。例如，2018年彭博社報導稱，中國特工在Supermicro主機板中植入間諜晶片，但該報導遭到相關公司的強烈否認。



據日前新聞報導，美國司法部在加州逮補了兩名中國籍人士，指控後者涉嫌非法出口包括輝達H100高階AI晶片在內，價值數千萬美元的高階AI晶片到中國，兩人已遭到逮捕並起訴

「後門」風暴背後的美中角力

有分析稱，AI晶片製造與海外銷售通常才是重點，但在如當前北京與華盛頓之間這種脆弱、缺乏信任的科技及關稅大戰的局勢下，便成為政治問題。

H20本是輝達為中國市場設計的晶片，在上個月解禁。H20的性能估計僅有輝達另一款高階AI晶片H100的70%左右，但已是輝達獲准在中國銷售性能最強大的AI晶片。

因此，中國網信辦的聲明，將H20或輝達在中國銷售之路再次設下障礙，即便北京才剛高調歡迎黃仁勳如搖滾明星般的在今年三度拜訪北京。專家向BBC說，這一事件不僅突顯了晶片安全的技術挑戰，也暴露了中美之間在科技領域日益緊張的關係，以及中國不願依賴西方高科技的高調立場及姿態展現。

根據路透社報導，H20近日已經先收到30萬筆訂單，但隨後傳出網信辦以安全疑慮問題約談輝達。有分析稱這反映了中國市場（特別是騰訊及阿里巴巴等大廠）對輝達AI晶片的強烈需求。但是，市場之上的政府對外國科技的忌憚仍大。因為中國在科技自主化道路上的主導戰略及自信，在這些年來水漲船高。換言之，研發國產晶片，減少對美國及西方技術的依賴成了主旋律。此次輝達H20晶片銷往中國的一波三折背後，有分析師甚至認為H20晶片，甚至輝達對中國來說已經「可有可無」，因為國產華為生產的晶片算力可以與其匹敵。

對此，半導體分析師，美國科技諮詢公司The Futurum Group研究部總監王韋傑（Ray Wang）向BBC中文解釋，網信辦發出的警告似乎比較針對中國國內的企業，特別是國有企業及涉及關鍵基礎設施的公司，而非針對輝達。換言之，中國的作法可能更多是為了向國內企業和公眾展示其對科技安全的重視，而非立即終止與輝達的合作。而這種策略在中美科技競爭加劇的背景下，兼具技術和政治意義。

「在商業應用中，後門可能是設計缺陷或疏忽的結果，但在軍事或關鍵基礎設施中，後門可能被視為國家安全的直接威脅。」王韋傑說。

不過，王先生仍強調，就他的研究來看，H20與輝達公司對中國仍至關重要，原因之一是相較於中國本土晶片製造商的產品來看，輝達的H20仍是中國大廠（如騰訊、百度、字節跳動、阿里巴巴）以及頂尖AI初創企業（如月之暗面、DeepSeek等）的首選。王韋傑又向BBC說，H20的算力和軟體生態（如CUDA平台）為中國AI產業提供了無可替代的價值，尤其在訓練大型語言模型和高效能運算應用中。

從這一角度講，輝達H20晶片的後門疑慮不僅是一個技術問題，更反映了中美在科技主導權和國家安全上的激烈競爭。據路透社本月（5日）報導，美國司法部在加州逮補了兩名中國籍人士，指控後者涉嫌非法出口包括輝達H100高階AI晶片在內，價值數千萬美元的高階AI晶片到中國，兩人已遭到逮捕並起訴。日經新聞則在本週刊出報導，由輝達供貨的兩家新興汽車公司——小鵬汽車與蔚來汽車——正透過自主研發晶片，在其最新車型中使用了自研晶片：小鵬的「圖靈」（Turing）與蔚來的「神機NX9031」晶片，目的是降低對輝達晶片的依賴。

雖然輝達強調其安全措施，但中國的質疑表明，技術信任與地緣政治信任的脫節正在放大這一問題的影響。對中國市場而言，H20仍是大熱的產品，但北京的強硬立場也顯示了其科技自主化的決心。

H20晶片是什麼？

H20晶片英偉達專為中國市場開發的圖形處理單元 (GPU)，旨在遵守美國對先進人工智慧晶片的出口限制 它是英偉達高階H100晶片的弱化版本，降低了互連頻寬並進行了其他修改，以滿足美國工業和安全局設定的出口管制門檻。儘管有局限性 ，H20仍然是一款適用於特定推理工作負載的強大晶片，並被中國科技公司用於 AI 開發。

與 H100 相比，H20 的互連頻寬有所降低，這影響了其在需要處理器之間高速資料交換的 AI 訓練任務中的效能。

根據報道，雖然 H20 的記憶體（96GB HBM3）比 H100（80GB HBM3e）更大，但其記憶體頻寬較低（4.0 TB/s 對 4.8 TB/s）。

H20 專為特定推理工作負載而設計，特別是那些需要高記憶體頻寬和批次效率的工作負載。