



Розвиток е-інфраструктури в системі освіти і науки України

Станіслав Зуб,

доктор фізико-математичних наук, професор,

Андрій Прокопенко,

доктор педагогічних наук, професор,

Ірина Яловега,

кандидат технічних наук, доцент,

Харківський національний педагогічний університет
імені Г. С. Сковороди

Враховуючи процеси глобалізації, для збереження та розвитку власної науки та освіти Україна має інтегруватися в Європейський дослідницький простір. Невід'ємною частиною цього простору є е-інфраструктура, яка задає певний рівень цифровізації наукового устаткування та систем, що з ним пов'язані. Таким чином, приєднання України до Європейського дослідницького простору потребує наявності досить розвиненої е-інфраструктури. І як буде показано в роботі, в існуючих умовах розвиток такої е-інфраструктури вимагає невідкладного її створення у системі освіти України. На жаль, на теперішній час спостерігається значне цифрове та технологічне відставання українських дослідницьких інфраструктур у порівнянні з розвинутими країнами. Фінансування освіти і науки за залишковим принципом не дає можливості очікувати значних інвестицій з боку держави, особливо в складних умовах сьогодення нашої країни. Формування ефективною е-інфраструктури, навіть при обмеженому державному фінансуванні, надасть можливість значно підвищити рівень як освітніх послуг, так і наукових досліджень.

Е-інфраструктура або цифрова інфраструктура потрібна там, де існує необхідність в дистанційному доступі до масивів даних, які, наприклад, використовуються в освітніх процесах (віртуальні лабораторії, тощо), або в науковій діяльності, де великі об'єми даних генерують потужні експериментальні установки. На жаль, нічого подібного поки в Україні не працює. Єдина сфера, яка оперує з великими даними і генерує в нашій країні попит на е-інфраструктуру, — телекомунікаційна. Проте, оскільки Україна сповідує принципи відкритості економіки та не захищає свій інформаційний простір і ринок від зовнішніх впливів, то навіть цей невеликий попит реалізується не за рахунок нашої науки та внутрішніх постачальників послуг, а шляхом зовнішнього запозичення, тобто імпорту. Свого часу так був знищений внутрішній ринок програмування — через відсутність законодавчої бази на інтелектуальний продукт піратські копії західного програмного продукту повністю нівелювали зусилля вітчизняних програмістів. Сьогодні це повторюється з нашими вченими, послуги яких не потрібні, оскільки є результати і напрацювання вчених інших країн. Це замкнуте коло з

позитивним зворотним зв'язком, тобто система пішла шкереберть. Рухаючись цим шляхом завтра у державі будуть знищені можливості розробки власного аутентичного продукту й у інших, більш соціальних, аніж прибуткових галузях: освіті і медицині, сільському господарстві.

Дослідницькі е-інфраструктури мають давню історію. Найбільше розповсюдження в Європі отримали три платформи ґрідів [1], які базуються на так званому сполучному програмному забезпеченні (grid middleware): gLite, ARC (Advanced Resource Connector), UNICORE (UNiform Interface to COmputing REsources) та dCache. Всі вони в тій чи іншій мірі беруть свій початок від проекту Globus Toolkit, перша версія якого з'явилась ще в 1998 р., а остання в 2018 р. Сучасні пакети grid middleware крім інструментальних засобів включають широкий набір служб. Так, наприклад, gLite [2] було створено в межах проекту EGEE (Enabling Grids for E-Science — "Розгортання ґрід для е-науки") на основі сполучного програмного забезпечення LCG (Large Hadron Collider Computing Grid) [3], призначеного для великого адронного колайдера, який був побудований у ЦЕРНі для вивчення фундаментальних властивостей субатомних частинок та сил. Більшість діючих на сьогодні ґрід-систем обробляють колосальні об'єми даних в десятки петабайт щороку та мають обчислювальні потужності в сотні тисяч процесорних ядер.



Функціонування таких е-інфраструктур потребує уніфікації засобів сполучного програмного забезпечення, або UMD (Unified Middleware Distribution) [4], до якого відносяться вищезгадані платформи gLite, ARC, UNICORE, dCache. Компоненти ґрідів можуть функціонувати в різних операційних системах та взаємодіяти завдяки стандарту Open Grid Forum (OGF). Ґрід-сервіс є веб-сервісом за стандартом OGSA (Open Grid Services Architecture). Архітектура OGSA визначає базову модель сервісів виконання (Basic Execution Service) та відповідає стандарту GLUE2, який описує

ресурси інформаційних сервісів для забезпечення їх інтегрованих інтероперабельності. Проект European Grid Infrastructure (EGI) [5, 6] створено для об'єднання зусиль проектів gLite, ARC та ряду інших для розробки уніфікованого набору компонентів. Отже, EGI — це об'єднана е-інфраструктура, створена для надання сучасних обчислювальних послуг в наукових дослідженнях та інноваціях. Е-інфраструктура EGI фінансується державним коштом та включає велику кількість центрів обробки даних та хмарних постачальників, розповсюджених в Європі та в усьому світі. В цьому проекті не передбачено створювати нове ПЗ для ґрід-систем, а лише інтегрувати в репозитарії UMD вже наявні або розроблені третіми сторонами програмні засоби, тестувати їх, та оцінювати продуктивність та якість.

Ми не ставимо на меті детально розглядати архітектуру зазначених проектів. Певною мірою це висвітлено в роботах на які ми посилаємось. Нашою метою є розгляд існуючої ситуації з розвитку е-інфраструктури в Україні, її проблеми та шляхи вирішення.

Як можна виправити ситуацію? Перш за все, потрібно відмовитися від тези, що ринок все сам владнає і відрегулює. В прогресивних країнах — це давно застаріла парадигма, але в Україні вона ще й досі вважається аксіомою. Тому необхідно включити важелі державного управління. Але для прозорого, компетентного, ефективного керування потрібна розвинена система моніторингу та аналізу економічної інформації, основною складовою якої і може бути е-інфраструктура. Через погане фінансування і нехтування потребами і розвитком освіти і науки, вони знаходяться у поганому стані, але ще функціонують. Так, система освіти генерує попит на телекомунікаційні послуги і дослідницьке устаткування для навчання.

Е-інфраструктура може в значній мірі підвищити рівень захисту даних в освітніх та наукових процесах. Досвід останніх років показує, що, наприклад, банківська система, котра використовує комерційні

системи безпеки, які не пройшли наукову експертизу, не здатна протистояти сучасним викликам. Система безпеки нагадує систему фортифікації — окрема будівля, навіть дуже захищена, не витримає навали ворога. Так само за архітектурним планом будують міста. Коли забудовники зневажають архітектурні плани, маємо екологічні катастрофи на зразок підтоплення вокзалу «Південний» у Києві у 2017 р. На жаль, безлад, що діється в сфері електронних комунікацій, не такий очевидний, як прорив каналізаційного колектора, але шкода для держави може бути набагато більша. Кібератаки завдали істотної шкоди іміджу та економіці держави. Це говорить про серйозні системні проблеми у захисті. За використання правильно організованої е-інфраструктури ці атаки були б швидко ідентифіковані та локалізовані, як це сталося в інших країнах, де подібні е-інфраструктури вже створені і функціонують. Так само, як і будівництво міст та фортифікацій, будівництво е-інфраструктур вимагає чіткого продуманого плану і кваліфікованої реалізації.

Концепція розвитку е-інфраструктури в Україні

1. Необхідно на законодавчому рівні зафіксувати пріоритет науки під час ухвалення експертних рішень в області **цифрового права, кібербезпеки** і розвитку **е-інфраструктури** в Україні.

2. Створити широке експертне товариство з представників МОН, НАН, НАКА і НАПН України для обговорення шляхів інтеграції і пріоритетних напрямів розвитку власної е-інфраструктури. Максимально інтегрувати наявні е-інфраструктури в європейські наукові та мережеві структури і проекти, які вже пройшли наукову експертизу і використовуються в Європі і США.

3. На базі МОН створити постійно діючі тренінгові центри для персоналу і користувачів е-інфраструктури із залученням вітчизняних і західних фахівців.

4. Створити фонд для підтримки наукових стартапів з розбудови е-інфраструктури. Передусім, для створення

програмних продуктів, спрямованих на реалізацію **стандартів вітчизняної е-інфраструктури** (мають бути розроблені), які отримали підтримку експертного наукового співтовариства.

5. Створити державний портал для висвітлення діяльності з розвитку е-інфраструктури та її прозорого фінансування на конкурсних засадах. На порталі розмістити базу ПЗ, яке визнане основою для розробки стартапів.

6. Необхідно на законодавчому рівні ввести монополію на **цифрові сертифікати та сервіси реєстрації віртуальних організацій** для роботи в е-інфраструктурі і підпорядкувати ці центри експертному співтовариству МОН і НАН України. Забезпечити максимальну автоматизацію і прозорість цього процесу. Виключити комерційний інтерес в цій діяльності. Важливо, щоб доступ в е-інфраструктуру мав уніфікований характер, що полегшить розробку ПЗ і підвищить безпеку системи. Система безпеки має ґрунтуватись на цифрових сертифікатах і віртуальних організаціях, як в ґрид.

Для розвитку е-інфраструктури вкрай важливо інтегрувати в неї датчики, прилади та інше складне устаткування. Це дозволить забезпечити безпечний дистанційний доступ для широкого кола споживачів. Тобто, дорогі прилади (мікроскопи, стенди і таке інше) можна використати дистанційно. При цьому організувати їх обслуговування максимально професійно, тоді віддача від їх використання буде максимальною. Розпочавши цей процес з науково-вимірювальних приладів, можна перейти до пристроїв, цікавих для комерційних структур. Тоді трансфер технологій буде здійснюватиметься з максимальною користю для освіти і науки.

Є очевидною тенденція переходу до тонких клієнтів (мобільні телефони, планшети, нетбуки). Необхідно запропонувати компаніям програмістів доступ до наукових ресурсних центрів. Щоб конкурувати з «MS Azure» та «Amazon», необхідно забезпечити пільговий режим роботи і максимальну якість та безпеку. Потрібно

зробити цей сектор вигідним для інвестицій.

Усі перелічені дії, включаючи монополізацію ряду функцій, спрямовані не на збір ренти з користувачів за рахунок виняткового стану, а на відновлення паритету із зовнішніми гравцями, який був втрачений через тривалу відсутність підтримки науки і освіти з боку держави. Ми вважаємо, що з часом стандарти та уніфікація процедур і функцій е-інфраструктурі дозволить відмовитися від ряду монопольних функцій.

7. Треба зняти обмеження, які накладені на наукові та освітні установи щодо розпорядження бюджетними коштами, які істотно гальмують розвиток наукової сфери. А саме:

ліквідувати обмеження щодо граничних сум на придбання обчислювальної техніки, програмних засобів, приладів, обладнання та устаткування, що безпосередньо використовується в науковій та науково-технічній діяльності (постанова КМУ від 4 квітня 2001 р. № 332 та інші нормативні акти);

«Про внесення змін до Порядку реєстрації технологій та їх складових, що створені чи придбані за бюджетні кошти або створені чи придбані підприємствами державної форми власності» у частині можливості виключення з процедури підприємства державної форми власності, які не фінансуються державою та не мають за потрібне здійснювати трансфер технологій.

8. Розробити спеціальний механізм фінансування наукових та науково-технічних розробок в галузі національної безпеки (е-інфраструктура, оборона, ядерні та космічні технології).

9. Внести зміни у принципи формування державних витрат на науку, а саме — видатки державного бюджету на науково-дослідні або науково-технічні роботи мають бути скеровані за такими напрямками:

базове бюджетне фінансування наукових установ, що складається із коштів на утримання інфраструктури (включаючи

е-інфраструктуру), обладнання наукових установ та науковців. Це фінансування спрямовується головним розпорядникам бюджетних коштів напряду.

10. Забезпечити виконання положень VI Прикінцеві та перехідні положення Закону України «Про наукову і науково-технічну діяльність», що набувають чинності з 1 січня 2020 року, а саме:

ч.2 статті 36: «... Держава гарантує встановлення ставок (окладів науковим працівникам державних наукових установ та вищих навчальних закладів), виходячи з розрахунку посадового окладу м.н.с. на рівні не нижче подвійної середньої заробітної плати у промисловості України»;

ч.2 статті 36: «... Держава забезпечує бюджетне фінансування наукової і науково-технічної діяльності у розмірі не менше 1,7 відсотка ВВП України».

Комерційні перспективи

1. Створення зручного та безпечного середовища для співтовариства програмістів. Це середовище має мати якісне устаткування, низькі ціни на ресурси, детальну документацію, високу якість зовнішньої та внутрішньої безпеки та гнучку демократичну структуру управління. Це зробить е-інфраструктуру привабливою для інвестицій.

2. Створення ПЗ, яке дозволяє інтегрувати наукове устаткування в е-інфраструктуру, дуже важливо для науки, освіти і бізнесу. Пізніше це ПЗ може бути також застосовано для бізнесу (збір даних з лічильників та моніторингу і таке інше).

3. Уніфікована і стандартизована система для усіх секторів економіки законодавчо захищена від свавілля силових структур. Свого роду вільна економічна зона з прозорими і пільговими правилами для освітніх закладів, наукових проектів, IT-бізнесу та фрілансерів.

Інтеграція наукового устаткування в е-інфраструктуру

Інтеграція наукового устаткування в е-інфраструктуру дозволить ефективно використати дороге устаткування **Центрів загального користування** і дозволить

з'єднати устаткування з його споживачами в географічно рознесених регіонах, а також забезпечить високу міру довіри до даних вимірювань, що передаються через глобальні мережі.

Координація і управління процесами вимірів, часто розділених великими відстанями, вимагає підтримки з боку сучасних інформаційних технологій. Вважається, що така підтримка може бути основана на технологіях інтернету. Проте, в цій системі не передбачені можливості управління вимірювальними і обчислювальними процесами, а дані, які передаються постійно, знаходяться під загрозою фальсифікації і, навіть, безповоротної втрати. Проблеми взаємодії віддалених приладів, а також захисту даних вимірювань в глобальних мережах вирішуються, але ці рішення завжди часткові, пристосовані до конкретного завдання і вимагають значних спеціальних зусиль.

На тому ж устаткуванні (тобто на тих самих елементах телекомунікацій, які використовує інтернет) може функціонувати якісно нова система глобальних розподілених обчислень, що забезпечує найвищий можливий на сьогодні рівень безпеки, який виключає навіть такі хворобливі явища, як віруси і хакерські атаки. Нова парадигма розподілених обчислень дістала назву грид. Ця система стала відповіддю на виклики, які пов'язані з якісним стрибком підвищення складності експериментальних фізичних установок та інформаційних систем, які вимагають нових підходів до моніторингу, управління та обслуговування.

Існує безліч реалізацій цієї концепції в США та Євросоюзі. Впродовж останніх десяти років, як вже було зазначено, особливе місце серед реалізацій грид займають розробки проектів EGEE та EGI. На даний момент ці розробки входять до пакету Unified Middleware Distribution (UMD). Річ у тому, що саме UMD (спадкоємець gLite Middleware) використовується для обслуговування найскладнішої фізичної (та вимірювальної) установки, яка створена людством, — Large Hadron Collider

(LHC) в ЦЕРН — Європейській раді з ядерних досліджень, де вперше зіткнулися з проблемою нестачі обчислювальних ресурсів для обробки колосального об'єму інформації, що народжується в експериментах на LHC. UMD створювалося і тестувалося унікальною командою найбільш висококваліфікованих програмістів і фізиків для застосування у виключно напружених умовах. Зібрати настільки сильну команду для вирішення такого грандіозного завдання в найближчому майбутньому навряд чи буде можливо. Незважаючи на таку спрямованість, UMD розроблялося як універсальна платформа для створення розподіленої обчислювальної мережі, яка повинна об'єднати National Grid Infrastructures (NGI) у рамках проекту EGI. Проект наслідував розробки EGEE за фінансової підтримки Євросоюзу. Зараз UMD набуває широкого поширення не лише для завдань фізики високих енергій, але й для вирішення трудомістких завдань біології, медицини і в інших галузях науки. Як правило, грид-розробки спрямовані на розподіл обчислювальних ресурсів і ресурсів пам'яті. І дійсно, більшість програм і користувачів потребують саме цих ресурсів. Але слід підкреслити, що віддалене управління та збір даних із самого початку є частиною існуючої концепції грид так само, як і струнка та ефективна система захисту даних.

Безпека лежить в основі грид. Її складовими є як спеціальні протоколи безпечної передачі даних такі, як GSIFTP, Secure RFIO, Gsidcap, так і система авторизації, яка у свою чергу є багаторівневою і забезпечує глибоке розмежування прав доступу до ресурсів грид. GSIFTP надає функціональність FTP-протоколу, але з підтримкою Grid Security Infrastructure (GSI). GSI забезпечує безпечну роботу в незахищених мережах загального доступу, надаючи такі сервіси, як аутентифікація, конфіденційність передачі інформації і уніфікований вхід в грид.

В якості ідентифікаторів користувачів і ресурсів в GSI використовуються цифр-

рові сертифікати X.509. Цей протокол є відповідальним за швидку, безпечну і ефективну передачу файлів. Він надає як можливість управляти передачею файлів між двома віддаленими по відношенню до клієнта елементами зберігання даних (third-party transfers), так і передавати дані кількома паралельними потоками.

Nовим є використання VOMS-сервісу (Virtual Organisation Membership Service), що зберігає інформацію про приналежність користувачів до певних груп і віртуальних організацій та про роль в них. Права користувача можуть встановлюватися залежно від його приналежності до тієї або іншої віртуальної організації, групи та ролі (за допомогою VOMS). Таким чином, система безпеки грід має усі відомі на сьогодні засоби криптографії та оснащена стрункою системою авторизації та розділення прав, яка заснована на цифрових сертифікатах.

Як вже було сказано, LHC є найскладнішою не лише фізичною, але і вимірювальною установкою, створеною людством. Здавалося б, на цьому прикладі вирішені всі завдання по зв'язку вимірювальних приладів і грід. Але це не зовсім так. grid middleware створювалося паралельно з LHC, і обробка даних з детекторів LHC є не лише програмним, але і апаратно-програмним рішенням, доступним тільки розробникам. Сьогодні все частіше «рядові» користувачі потребують інтеграції вимірювальних приладів з грід для побудови інформаційно-вимірювальних систем з підвищеним захистом даних, що передаються на значні відстані, а також для дистанційного доступу до приладів, в принципі, з будь-якого сайту грід. Для вирішення цих завдань було створено проєкт Grid Enabled Remote Instrumentation with Distributed Control and Computation (GridCC). Метою цього проєкту було застосування можливостей грід для безпечної, віддаленої, спільної роботи команди з моніторингу та управління приладами та даними, які були згенеровані та зберігаються на розподіленому науковому устаткуванні, використовуючи традиційні

ресурси грід. У рамках цього проєкту був розроблений Instrument Element (IE) [7], який успішно застосовується різними науковими колабораціями для віддаленої взаємодії з приладами в середовищі грід.

IE складається з пов'язаної колекції сервісів, котрі і забезпечують усю функціональність для конфігурації, управління і моніторингу вимірювальними приладами, які знаходяться за IE-інтерфейсом, що забезпечує їх взаємодію з іншим грідом. Перелічимо важливі характеристики IE. Він забезпечує наступні функціональні вимоги: уніфіковану модель вимірювального приладу; стандартний грід-доступ до приладів; можливість взаємодії між різними приладами, що належать різним інститутам і віртуальним організаціям (VO). Користувачем IE може бути не лише людина, але і будь-яка програма, яка має певні права, які визначаються електронним грід-сертифікатом.

Користувач IE може мати одну з наступних ролей:

- спостерігач — право на моніторинг приладу;
- оператор — доступ до конфігурації та управління приладом і його моніторингу;
- адміністратор — право створювати конфігурацію приладу, яка після цього стає доступна спостерігачу та оператору.

У будь-який момент може бути безліч спостерігачів і адміністраторів, але тільки один оператор, який працює (тобто управляє) приладом. Якщо прилад складний і має незалежні модулі (наприклад, детектор CMS LHC), тоді декілька операторів можуть управляти кожен своїм модулем, для чого і призначено віртуальний диспетчерський пункт (VCR).

Прилад або комплекс приладів може бути як дуже простим, так і дуже складним. У зв'язку з цим, при розробці абстрактної моделі приладу була введена наступна класифікація приладів:

- dummy instrument (прилад без управління, датчик для збору даних);

- smart instrument (інтелектуальний, забезпечений мікропроцесором, продуктивний, зі швидким каналом зв'язку, здатний діяти автономно);
- smart instrument in an adhoc network (те ж, що і попередній, у складі спеціалізованої мережі виконують єдине вимірювальне завдання).

Сучасні технології у багатьох випадках дозволяють дистанційно порівнювати вимірювальні прилади. Існують міжнародні програми в області синхронізації годинника, навігації та в інших областях. Ґрід здатний вивести ці підходи на абсолютно новий якісний рівень та забезпечити більш високу продуктивність обробки і одночасний захист даних.

Складність ПО вимірювальних приладів зі вбудованою ОС постійно зростає. Все більше застосування знаходять вимірювальні прилади зі вбудованою ОС Linux. Як засоби виміру вони підпадають під вимоги законодавчої метрології. Складність тестування таких систем і проблеми захисту вимірювальних даних стають дуже актуальними. Мають бути виконані вказані вище вимоги L, T, D, що для таких систем представляє нове і непросте завдання. Проте, подібні завдання не є новими для розробників ґрід. На сьогодні, ЛНС є по суті, найскладнішим в історії цивілізації вимірювальним приладом, для якого ці та багато інших завдань вищого рівня складності вирішуються за допомогою ґрід. Наявні в ґрід інструменти разом з ІЕ надають усі можливості для ефективного і безпечного застосування вимірювальних приладів із вбудованою ОС Linux. В якості іншого підходу до рішення вимірювальних завдань з використанням складного ПЗ ми пропонуємо використати концепцію «тонких клієнтів». Інакше кажучи, прилади повинні мати мінімум ПЗ, що забезпечує інтеграцію в систему ґрід, а складна обробка «сирих» даних має виконуватись в середовищі ґрід. Вибір підходу залишається за розробниками вимірювального приладу.

Таким чином, освітня е-інфраструктура значно підвищує рівень технічних

можливостей та конкурентоспроможність нашої освіти, наприклад, в дистанційній формі навчання, яка на сьогодні вже є невід'ємною частиною навчального процесу багатьох освітніх закладів, а також може стати значним кроком до підвищення рівня освітніх послуг та розвитку науки в освіті, розробки інноваційних продуктів, комерціалізації досліджень в Україні.

Література

1. Foster I., Kesselman C. (Eds.). The Grid: Blueprint for a New Computing Infrastructure. Morgan — Kaufman, 1998. 701 p.
2. Grid middleware configuration at the KIPT CMS Linux cluster / S. Zub, L. Levchuk, P. Sorokin, D. Soroka // Nuclear Instruments and Methods in Physics Research A 559. 2006. P. 35–37.
3. LCG Middleware at the KIPT CMS Linux Cluster / M. V. Voronko, S. S. Zub, L. G. Levchuk, D. V. Soroka // Вісник Харківського національного університету. Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління». 703. 2005. С. 74–86.
4. Ґрід-технології в науці та освіті / С. С. Зуб, А. І. Прокопенко // Інформаційно-цифровий освітній простір України: трансформаційні процеси і перспективи розвитку. Матеріали методологічного семінару НАПН України. 4 квітня, 2019 р. С. 26–29.
5. Цифрова наука в программе «Горизонт 2020» / М. З. Згуровський, А. І. Петренко // Системні дослідження та інформаційні технології. 2015. № 1. С. 7–20.
6. The EGI Federated Cloud e-Infrastructure / Enol Fernandez-del-Castillo, Diego Scardacia, Alvaro Lopez Garcia // Procedia Computer Science. 68. 2015. P. 196 — 205.
7. The computer infrastructure grid: prospects for applications in metrology / P. I. Neyezhmakov, S. I. Zub, S. S. Zub // OIIML Bulletin. Vol LIII. № 3. 2012. P. 5 — 10.

05.01.2020