The Quantum Computing Company™

Detecting Lateral Movement with a Compute-Intense Graph Kernel

<anonymous collaborator> <US government agency>

Steve Reinhardt D-Wave Government spr@dwavesys.com

We implemented a compute-intensive graph kernel that finds lateral-movement-like behavior in netflow data and can execute quantumly in part. We sketch the remaining work to deliver quantum acceleration from graph kernels. We believe this enables a valuable new set of tools for cyber analysts.

Agenda

➡ Detecting lateral movement via maximum independent set

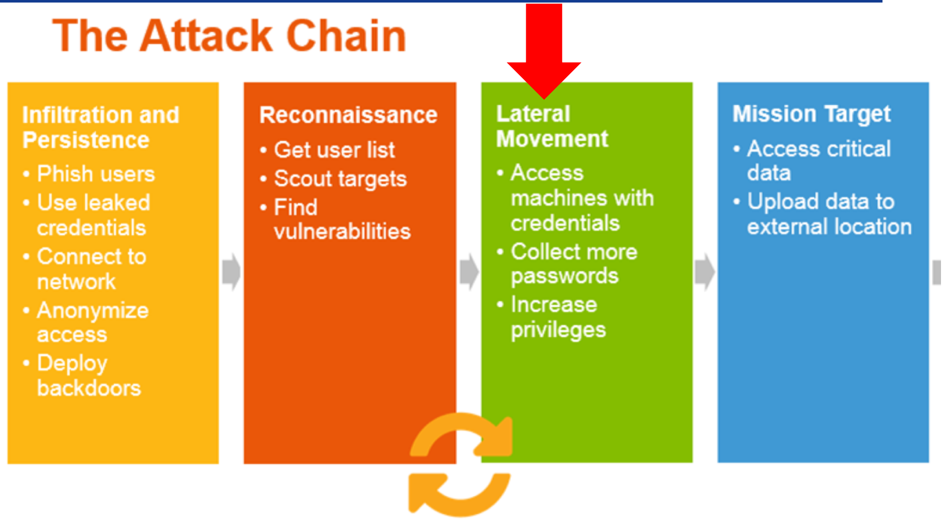
- Achieving high performance with graph kernels on a D-Wave system
- Implications for cyber and other analyses

Motivation

- Execution of some compute-intense graph kernels on D-Wave systems has yielded better answers than classical counterparts
  - Mniszewski et al., Quantum Annealing Approaches to Graph Partitioning on the D-Wave System, https://dwavefederal.com/app/uploads/2017/10/Qubits-Day-2-Morning-4_Susan_LANL.pdf
- DWave_NetworkX includes a set of compute-intensive kernels
  - Minimum vertex cover, minimum vertex coloring, maximum cut, maximum independent set, maximal matching, signed social network
  - https://github.com/dwavesystems/dwave_networkx
- Given exponential growth of computation with problem size, analysts have avoided these kernels, which are becoming tractable
- For what cyber problems are those (or similar) kernels useful?

Detecting Lateral Movement

- On graph of point-to-point logins (ssh, RDP), calculate maximum independent set (largest set of non-adjacent vertices)
- As point-to-point logins explore more of the enterprise network, MIS shrinks



The Attack Chain

Infiltration and Persistence
- Phish users
- Use leaked credentials
- Connect to network
- Anonymize access
- Deploy backdoors

Reconnaissance
- Get user list
- Scout targets
- Find vulnerabilities

Lateral Movement
- Access machines with credentials
- Collect more passwords
- Increase privileges

Mission Target
- Access critical data
- Upload data to external location

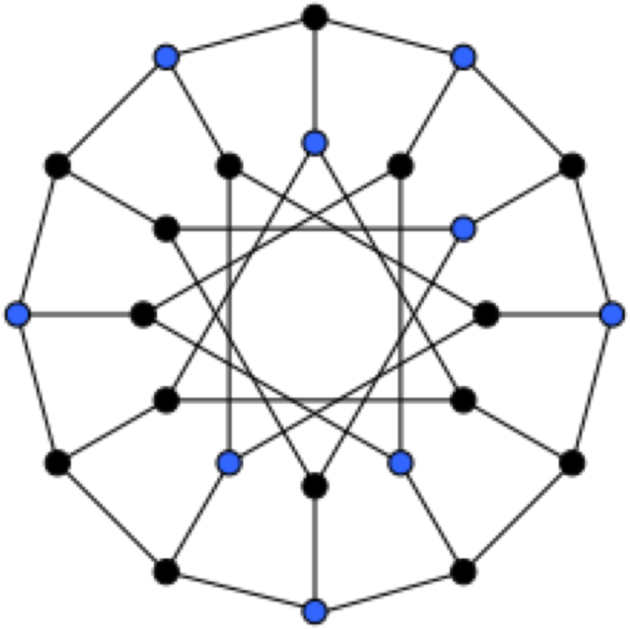
- [WIP] Confirmed on traffic not known to contain lateral movement; working to confirm on traffic with lateral movement

https://blog.rapid7.com/content/images/post-images/53326/the-attack-chain.png

D:Wave
The Quantum Computing Company™

Maximum Independent Set (MIS)

- An *independent set* is a set of vertices in a graph, no two of which are adjacent. A *maximum independent set* is an independent set of largest possible size for a given graph G.
- NP-hard problem



The nine blue vertices form a maximum independent set for the Generalized Petersen graph GP(12,4)

- **For exact solutions**, a set is independent if and only if its complement is a vertex cover. Therefore, the sum of the size of the largest independent set α(G) and the size of a minimum vertex cover β(G) is equal to the number of vertices in the graph.

https://en.wikipedia.org/wiki/Independent_set_(graph_theory)#Finding_maximum_independent_sets

Experiment

- LANL data, not known to contain lateral movement: https://csr.lanl.gov/data/2017.html
- 88 days of data; focused on first 8
- Used 4-day sliding time window
- Monitor shrinkage of max independent set as an indicator of lateral movement

Looking for a Good Graph Size

Good == relevant to large enterprise networks and feasible

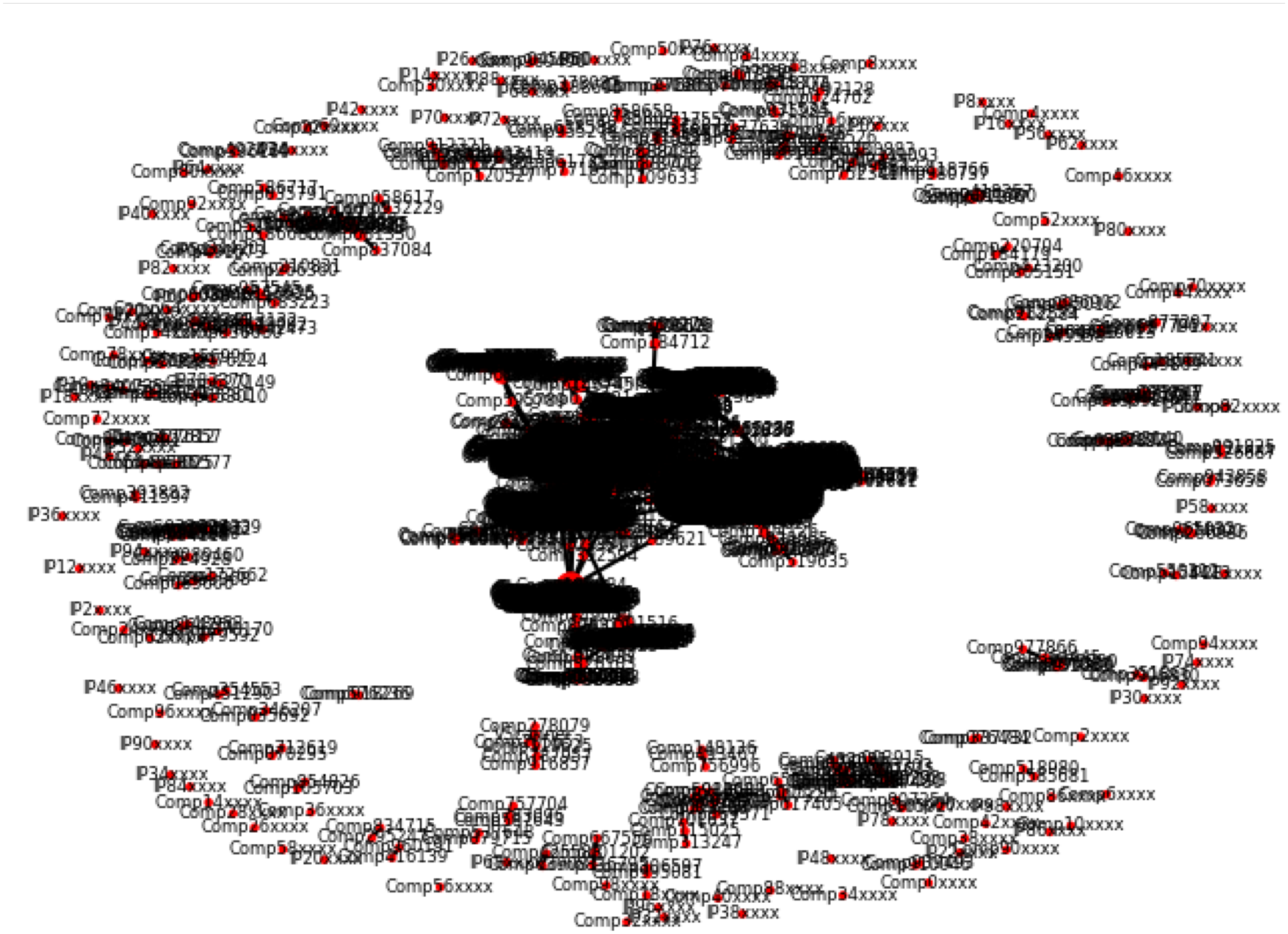
#IP addresses full (|V|)	#IP addresses reduced	#point-to-point pairs (|E|)	MIS size	time (s)
75571	1682	1474	1353	5.2
75571	3388	1474	3253	22.0
75571	21388	1474	21253	908.9
84718	84718	1239	?	> 18 hours

Averages of 5 timesteps except time-limit-exceeded
Running classically

Analytic Finds Smaller MIS Size

```
spr_mbp:10M $ /Users/sreinhardt/technical/app_code/cyberGraph/process3.py netflow netflow --nTimesteps 8  --nTimestepsPerDay 4 -
-inputFormat LANL --inputDigested --verbose --reduceGraph
Namespace(allInput='netflow', inputDigested=True, inputFormat='LANL', nRecordsPerTimeperiod=None, nTimesteps=8,
nTimestepsPerDay=4, nTimestepsPerWindow=None, pt2ptInput='netflow', reduceGraph=True, verbose=1)
Reducing graph? True
for timestep 3, the number of IP addresses in the full graph is 84718 and the number of IP-address-pairs that had point-to-point logins is
1239; the number of IP addresses in the reduced graph is 21178 and IP-address-pairs pt2tp is 1239
first MIS took 0:11:25.255169, second MIS took 0:17:05.962103
MIS size decreased from 21178 to 21059 (119) when ssh/telnet/RDP log-ins considered
  IP addresses that are no longer part of the maximum independent set are ['ActiveDirectory', 'Comp005825’, […]
for timestep 4, the number of IP addresses in the full graph is 82543 and the number of IP-address-pairs that had point-to-point logins is
1472; the number of IP addresses in the reduced graph is 21380 and IP-address-pairs pt2tp is 1472
first MIS took 0:13:12.690244, second MIS took 0:14:24.761931
MIS size decreased from 21380 to 21245 (135) when ssh/telnet/RDP log-ins considered
  IP addresses that are no longer part of the maximum independent set are ['ActiveDirectory', 'Comp005825’, […]
for timestep 5, the number of IP addresses in the full graph is 82266 and the number of IP-address-pairs that had point-to-point logins is
1525; the number of IP addresses in the reduced graph is 21427 and IP-address-pairs pt2tp is 1525
first MIS took 0:11:10.267566, second MIS took 0:11:31.594414
MIS size decreased from 21427 to 21294 (133) when ssh/telnet/RDP log-ins considered
  IP addresses that are no longer part of the maximum independent set are ['ActiveDirectory', 'Comp005295’, […]
for timestep 6, the number of IP addresses in the full graph is 66069 and the number of IP-address-pairs that had point-to-point logins is
1551; the number of IP addresses in the reduced graph is 21459 and IP-address-pairs pt2tp is 1551
first MIS took 0:11:06.538207, second MIS took 0:17:56.528730
MIS size decreased from 21459 to 21317 (142) when ssh/telnet/RDP log-ins considered
  IP addresses that are no longer part of the maximum independent set are ['ActiveDirectory', 'Comp005295’, […]
```

Preliminary Visualization



Discussion

- Why not use connected components (much faster)?
  - Could, but doesn’t give intuition about “how close to connected”
- Need to find Goldilocks-size graphs
  - If too small, runs fast enough classically
  - If too big, even with medium-scale quantum acceleration, exponential algorithm still infeasible
    - If QPU solves 10% of problem, 2^{10} combinations of those may be feasible
    - If QPU solves 1% of problem, 2^{100} combinations is not feasible

Next Steps

- Need to verify on data known to have lateral movement
- Need to find scale-appropriate viz that illustrates growth of connectedness (== shrinkage of MIS) for cyber analyst
- Code available via private Github repository
- Looking for collaborators, esp. with data

Agenda

- Detecting lateral movement via maximum independent set
- ➡ Achieving high performance with graph kernels on a D-Wave system
- Implications for cyber and other analyses

Two Main Paths to Quantum Computing

Gate-model architecture

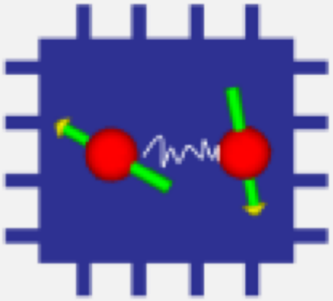
- Significant theoretical work since 1985, key algs defined in 1990s
- Major issue of error correction identified by Preskill in 1998
  - Believed to require 100-1000 physical qubits for every logical qubit
- Google recently announced system with 72 physical qubits, results TBD
- Digital nature in question
  - Preskill: “noisy intermediate-scale quantum” (NISQ) computers
- Current focus on approximate optimization algs (e.g., QAOA)

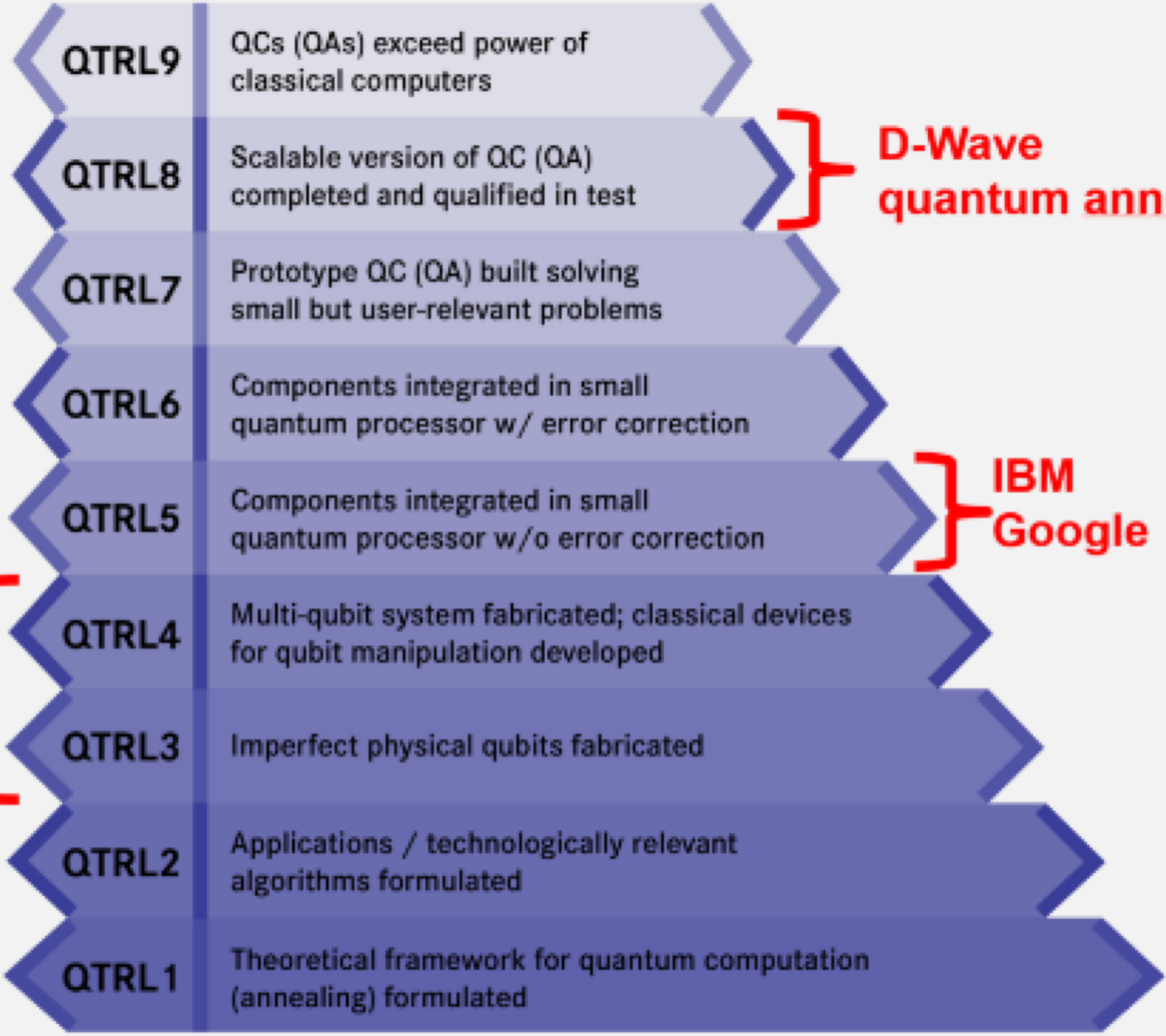
Quantum-annealing architecture

- Nishimori (1998) and Farhi (1999) described theory to find low energy states; Rose (2004) identified path to build such systems
- D-Wave (2010+) has delivered 4 generations of systems, the latest with 2000 qubits
- Academic knowledge is mostly empirical
- Problems friendly to D-Wave topology show ~1000X advantage; real-world problems ~parity
- New system generations every ~2yr

QTRL

Quantum Technology Readiness Levels describing the maturity of Quantum Computing Technology





QTRL9 QCs (QAs) exceed power of classical computers

QTRL8 Scalable version of QC (QA) completed and qualified in test

QTRL7 Prototype QC (QA) built solving small but user-relevant problems

QTRL6 Components integrated in small quantum processor w/ error correction

QTRL5 Components integrated in small quantum processor w/o error correction

QTRL4 Multi-qubit system fabricated; classical devices for qubit manipulation developed

QTRL3 Imperfect physical qubits fabricated

QTRL2 Applications / technologically relevant algorithms formulated

QTRL1 Theoretical framework for quantum computation (annealing) formulated

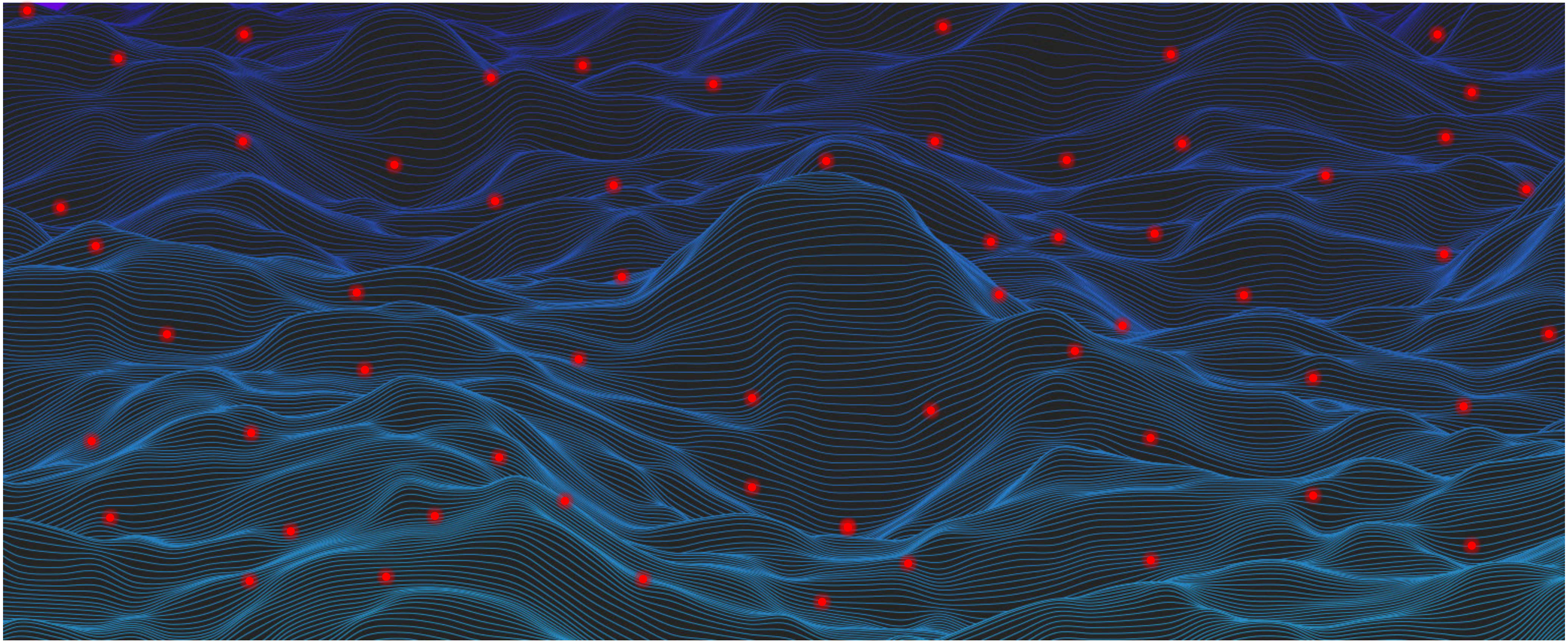
D-Wave quantum annealer

IBM
Google

Experimental qubit devices

© Kristel Michielsen, Thomas Lippert – Forschungszentrum Jülich (http://www.fz-juelich.de/ias/jsc/EN/Research/ModellingSimulation/QIP/QTRL/_node.html)

Quantum Annealing: How a D-Wave system works



Programming Model / Quantum Machine Instruction

QUBIT	q_i	Quantum bit which participates in annealing cycle and settles into one of two possible final states: {0,1}
COUPLER	$q_i q_j$	Physical device that allows one **qubit** to influence another **qubit**
WEIGHT	b_{ii}	Real-valued constant associated with each **qubit**, which influences the **qubit's** tendency to collapse into each of its two possible final states; **controlled by the programmer**
STRENGTH	b_{ij}	Real-valued constant associated with each **coupler**, which controls the influence exerted by one **qubit** on another; **controlled by the programmer**
OBJECTIVE	Obj	Real-valued function that is **minimized** during the annealing cycle

Known as

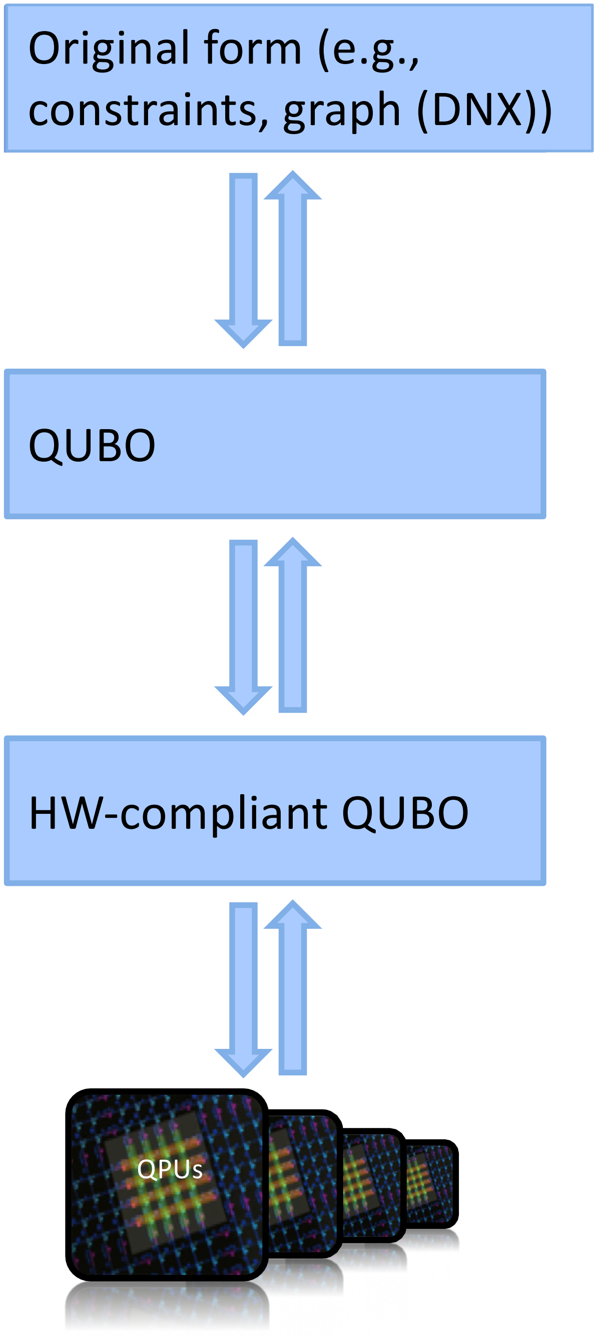
- Quadratic unconstrained binary optimization (QUBO) problem
- Ising model
- Unconstrained binary quadratic problem (UBQP)
- Probabilistic graphical model (PGM)

$$\boldsymbol{Obj(b_{ij};q_i) = \sum_{ij} b_{ij} q_i\, q_j}$$

The system **samples** from the q_i that minimize the objective

Note: The D-Wave 2000Q™ system added reverse annealing, which is a variant of this.

Mapping a Problem to D-Wave

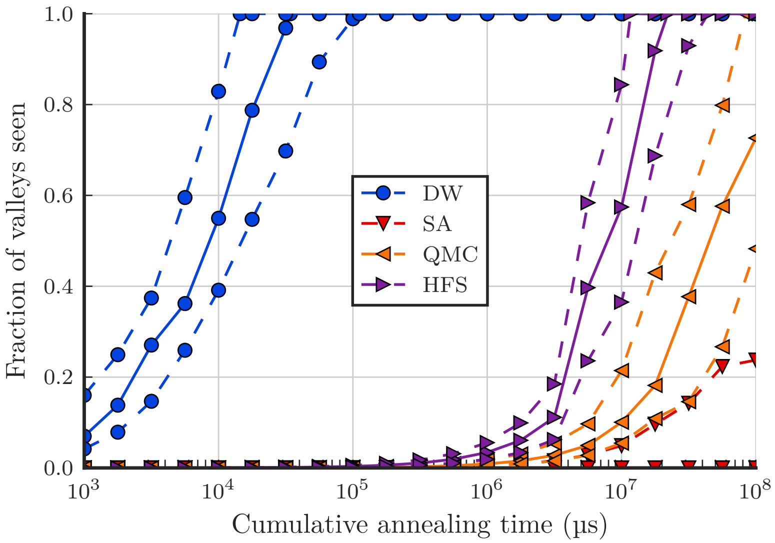


Original form (e.g., constraints, graph (DNX))

QUBO

HW-compliant QUBO

Steps	Best known methods
• Reduce problem size	
• Map to QUBO form	Avoid $O(N^2)$ #var growth
• Reduce QUBO	Find frozen variables
• Decompose QUBO	Various: energy impact, recomb elite
• Embed into HW graph	Avoid long chains, extend pre-embed
• Tolerate low precision	Bin values into discrete ranges



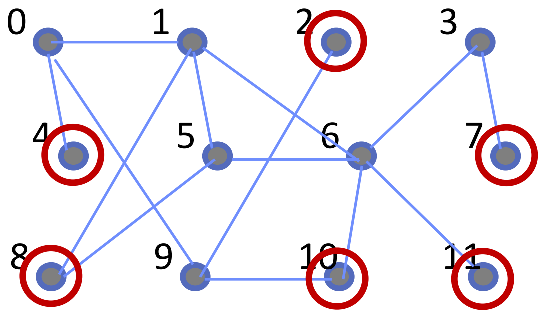
ılgebraic setting

s)

ıs as f(length)

The Quantum Computing Company™

Mapping MIS to QUBO Form



$$b_{i,j} \begin{cases} -1, & \text{if } i = j \\ 3, & \text{if } i < j \text{ and } ij \in E \\ 0, & \text{otherwise} \end{cases}$$

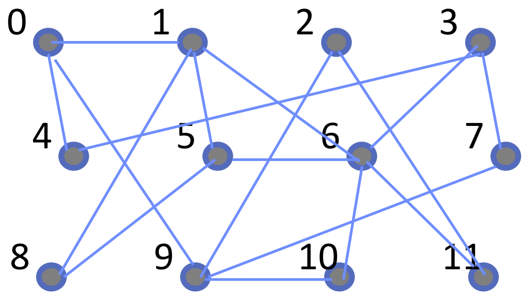
$$\begin{array}{c|cccccccccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
\hline
0 & -1 & 3 & & & 3 & & & & & 3 & & \\
1 & & -1 & & & & 3 & 3 & & 3 & & & \\
2 & & & -1 & & & & & & & 3 & & \\
3 & & & & -1 & & & 3 & 3 & & & & \\
4 & & & & & -1 & & & & & & & \\
5 & & & & & & -1 & 3 & & 3 & & & \\
6 & & & & & & & -1 & & & & 3 & 3 \\
7 & & & & & & & & -1 & & & & \\
8 & & & & & & & & & -1 & & & \\
9 & & & & & & & & & & -1 & 3 & \\
10 & & & & & & & & & & & -1 & \\
11 & & & & & & & & & & & & -1
\end{array}$$

$$\boldsymbol{Obj(b_{ij}; q_i) = \sum_{ij} b_{ij} q_i \, q_j}$$

Choose q_i that minimize

https://canvas.auckland.ac.nz/courses/14782/files/563551/download?verifier=mGDAXoF3TDAKoZfnjmQ1WBRfjlB0LkTc62wAdyGO&wrap=1

Mapping Traveling Salesperson to QUBO Form



- Many formulations use “for vertex i at step k” approach, which is O(N^2)

$$\boldsymbol{Obj(b_{ij}; q_i) = \sum_{ij} b_{ij} q_i \, q_j}$$

Choose q_i that minimize

Explicitly Hybrid Quantum/Classical Algorithms

- Due to Chapuis et al.
- CPU/GPUs and QPUs have drastically different natures; use each for its strengths
- Use classical techniques (k-core graph and core/halo partitioning) to partition graph into subgraphs that will fit in QPU, find max clique of each
  - Limited to finding cliques embeddable in the QPU

Chapuis, Djidjev, Hahn, and Rizk, "Finding Maximum Cliques on the D-Wave Quantum Annealer"

Cause for Optimism

NP-hard Problems Solved in Modern Compilers

Graph coloring	Code generation
Set-weighted covering	Register sufficiency
Topological sort	Instruction scheduling
Graph coloring	Register allocation, coalescing, minimizing spill, and reuse
Minimal vertex covering	Global reference allocation
Maximum weighted path cover	Array unification
Multiple graph partitioning	Distributed memory layout

Delivering Differentiated Performance

- Today (D-Wave 2000Q™): In practice, problems of **~64 variables** fit on the QPU
- Next-gen D-Wave system targeted at 4-5K qubits with denser topology

Aspect	Change	Effect on #variables in QMI	Notes
More qubits	2-2.5X more	* 1.4-1.6	
Denser topology	2.5X more	* 2.8	Higher perf due to shorter chains
QA changes	TBD		Lower noise, ◊
Better algs/tools		* 1.3	(e.g.) RBC embedding
Aggregate change		*** 5.46 == 326 vars**	

- Some problems shift from classically tractable to intractable between 64 and 326 variables: e.g., Markov networks (~50 today; 100s intractable)

◊ Roy et al.'s "Boosting integer factoring …" showed that per-qubit advance/delay of annealing in some cases led to a 1000X performance increase (i.e., fraction of valid results)

George

Agenda

- Detecting lateral movement via maximum independent set
- Achieving high performance with graph kernels on a D-Wave system
- ➡ Implications for cyber and other analyses

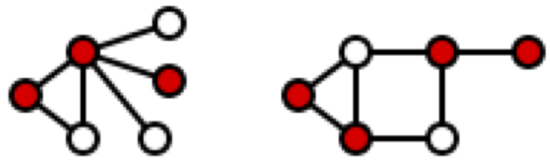
Detecting LM via MIS is Only One Use Case

- Current DWave_NetworkX kernels
  - Minimum vertex cover
  - Minimum vertex coloring
  - Maximum independent set
  - Maximum cut
  - Structural imbalance
  - Maximal matching

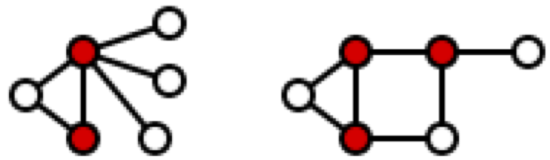
- N.B.: graph partitioning, community detection, and maximum clique implemented by LANL
  https://arxiv.org/pdf/1705.03082.pdf https://arxiv.org/pdf/1801.08649.pdf

Min Vertex Cover (MVCov)

- A *vertex cover V′* of an undirected graph *G* = (*V*,*E*) is
  a) a set of vertices where every edge has at least one endpoint in the vertex cover *V′*,
  b) a subset of *V* such that $uv \in E \Longrightarrow u \in V' \lor v \in V'$



- A *minimum vertex cover* is a vertex cover of smallest possible size (# vertices)



https://en.wikipedia.org/wiki/Vertex_cover

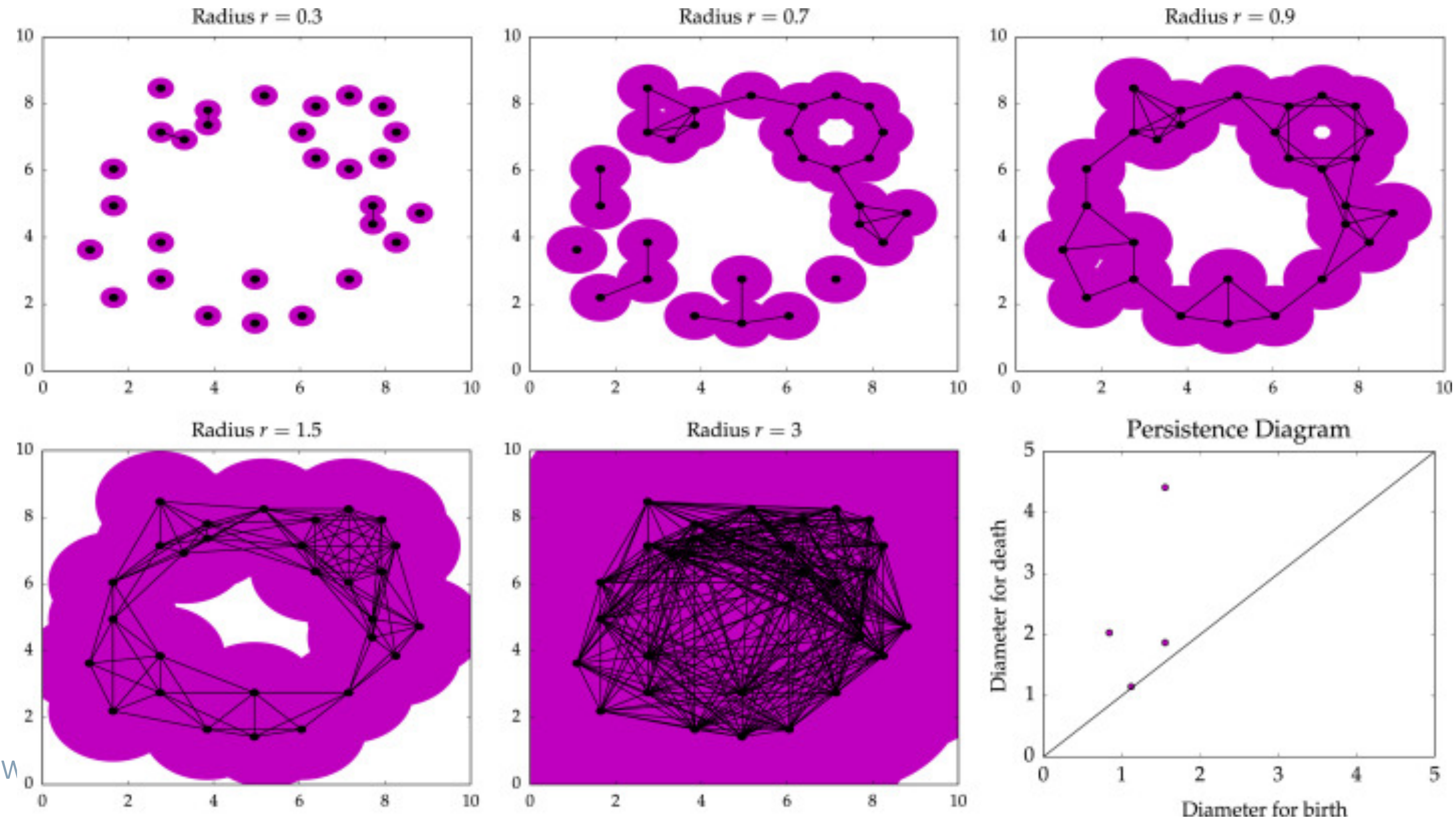
Cyber Use Cases

- 👍 In wireless sensor network, MVCov$^+$ equates to a plan for installing a patch that minimizes the #rounds while preserving full observability throughout
  https://infoscience.epfl.ch/record/225623/files/EPFL_TH7484.pdf
- 👍 MVCov is the optimal solution for worm propagation and hence for network defense
  Dharwadker and Pirzada, Applications of Graph Theory, ISBN 1466397098
- 👍 MVCov$^+$ finds minimal set of strings that occur in viruses but not in normal code https://math.mit.edu/~goemans/18434S06/setcover-tamara.pdf
- ★ To make the internet more robust, want better understanding of AS-level topology, but currently BGP data is being gathered from an incomplete and biased subset of ASes. MVCov$^+$ calculates the minimum set of watching ASes that can supply data for a comprehensive view of the internet.
  https://isolario.it/extra/publications/papers/BGPIncompleteness.pdf

https://www.flaticon.com/free-icon/thumb-up_13979#term=thumbs up&page=1&position=86

Related Use Case: Topological Data Analysis (TDA)

- E. Munch: “Find and quantify structure in noisy, complex data.”
- TDA’s Mapper capability commercialized by Ayasdi
- *Persistent homology* capability highlights the "relative prominence of homological features in the data set”



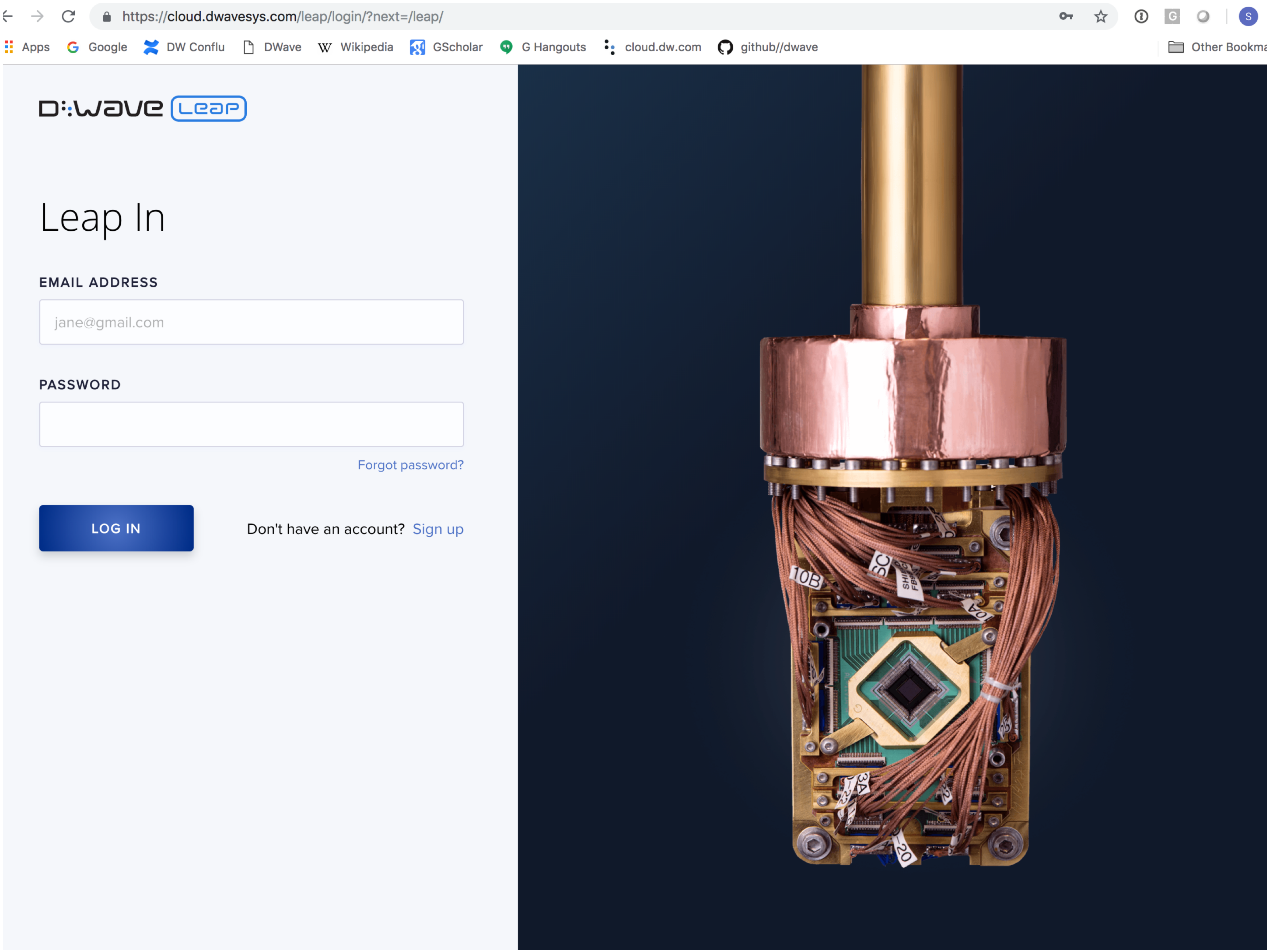
The Quantum Computing Company™

TDA / Persistent Homology (cont.)

- Believed to have high analytic value
  - Can serve to identify features on which machine learning learns
- Core kernels include:
  - Wasserstein (earthmover) distance between two distributions
  - minimum clique cover
- To date, min clique cover has been so expensive to compute that #dimensions analyzed is sharply limited, reducing analytic value
- D-Wave implementation of Wasserstein distance by Berwald et al.
- Exploring open-source release

J. Berwald, J. Gottlieb, E. Munch. Quantum Computation in a Topological Data Analysis Pipeline, https://www.dwavesys.com/sites/default/files/10_Tues_PM_TopPipe.pdf

We implemented a compute-intensive graph kernel that finds lateral-movement-like behavior in netflow data and can execute quantumly in part. We sketch the remaining work to deliver quantum acceleration from graph kernels. We believe this enables a valuable new set of tools for cyber analysts.



Leap In

EMAIL ADDRESS

jane@gmail.com

PASSWORD

Don't have an account? Sign up

For More Information, See

D-Wave Users Group Presentations:

- 2018 (N.America): https://www.dwavesys.com/qubits-north-america-2018
- 2018 (European): https://www.dwavesys.com/qubits-europe-2018
- 2017: **http://dwavefederal.com/qubits-2017/**
- 2016: https://dl.dropboxusercontent.com/u/127187/User%20Group%20Presentations-selected/Qubits_User_Group_Presentations_Index.html

LANL Rapid Response Projects:

- http://www.lanl.gov/projects//national-security-education-center/information-science-technology/dwave/index.php