Who Ya Gonna Call?
Global Health Outbreak, and Surveillance Technology

Seth Sims

System Admin/HPC Developer

August, 2016

Division of Viral Hepatitis

GHOST
Global Hepatitis Outbreak Surveillance Technology

Background Information

Haunted by Hepatitis

Background

- HCV infects approximately 3.5 million persons the United States and 180 million persons worldwide.
- Difficult to detect and investigate, because the infections can remain occult in >70% of infected persons for years, or decades.
- Risk of transmission is associated with exposure to blood and blood products
  - High risk groups: injection drug users and men-who-have-sex-with-men
- During 2005–2013 in the United States, 18 healthcare-associated outbreaks were detected, involving 228 outbreak-associated cases and >92,550 at-risk persons notified for screening

Hospital tech who spread hep C via drug thefts gets 30 years

By Jennifer Brown
The Denver Post

POSTED: 02/25/2010 01:00:00 AM MST | UPDATED: 6 YEARS AGO

8 COMMENTS

A surgical tech who infected at least 18 hospital patients with hepatitis C by stealing liquid painkillers and leaving behind her dirty syringes was sentenced to 30 years in prison Wednesday after a tearful plea for mercy, saying: "I am not the monster that drugs made me be."

Faced with the choice between Kristen Diane Parker's appeal for "another chance" and her victims' emotional claims that what she did amounted to attempted murder, U.S. District Judge Robert Blackburn sent Parker to prison for longer than called for under sentencing guidelines.

The typical sentencing range for the 27-year-old's crimes is 20 to

Kristen Parker: The former hospital

According to the 2014 CDC Hepatitis surveillance statistics report:

“A more recent study using death certificate data from 2003-2013 showed that the number of deaths associated with HCV infection continued to increase since 2007, and in 2012, surpassed the combined number of deaths associated with 60 other nationally notifiable infectious conditions that are routinely reported to CDC.”

The Common Welfare is My Business

What is GHOST

Distill sequence information into Public Health Intelligence

GHOST is a system

- Expert model modules (a.k.a. Cyber-molecular Assay)
- It is a massively distributed sequencing machine
- Requires no knowledge of bioinformatic techniques to utilize
- Fuses wet-lab and analysis techniques

Ghost current capabilities

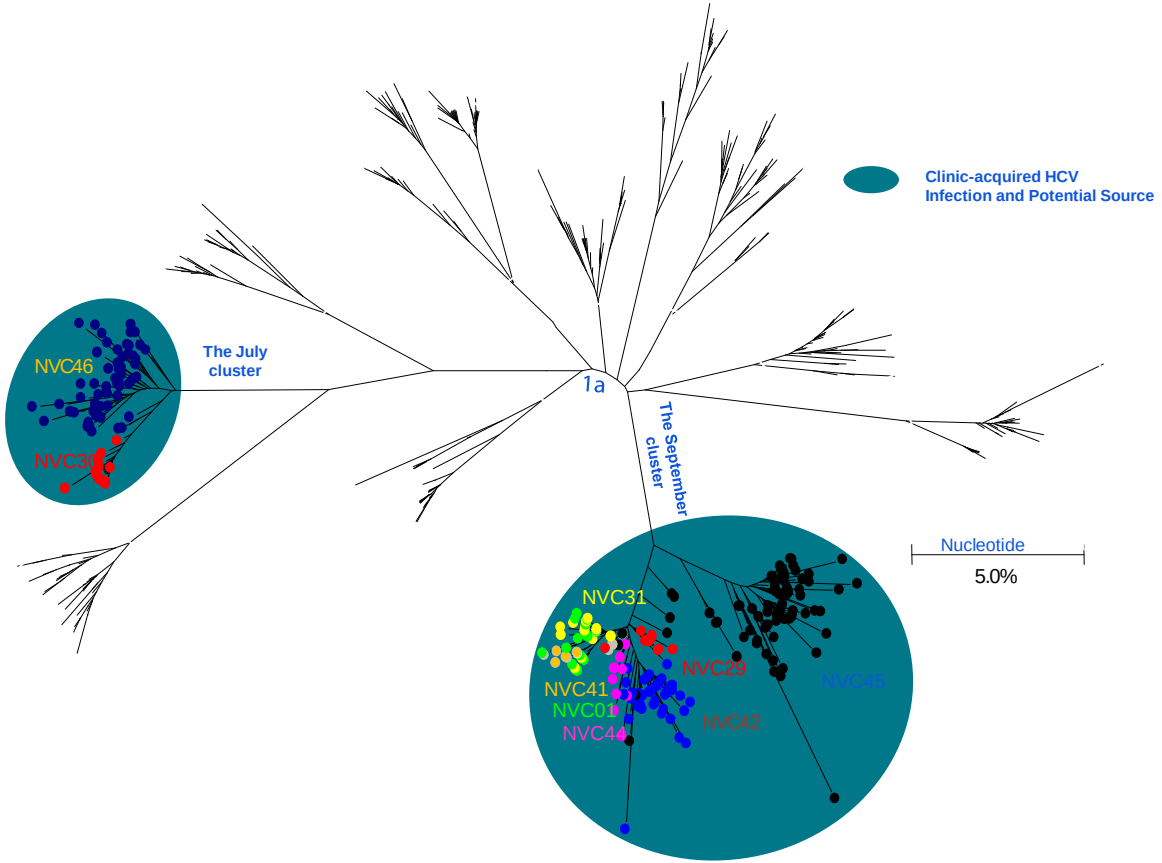
- GHOST's assays can currently take NGS samples and
  - Perform quality control
  - Assign genotype and calculate genetic heterogeneity
  - Detect transmission networks

The paper describing the methods was awarded the Charles C. Shepard Science Award, year 2016, in the "Laboratory and Methods" Category.

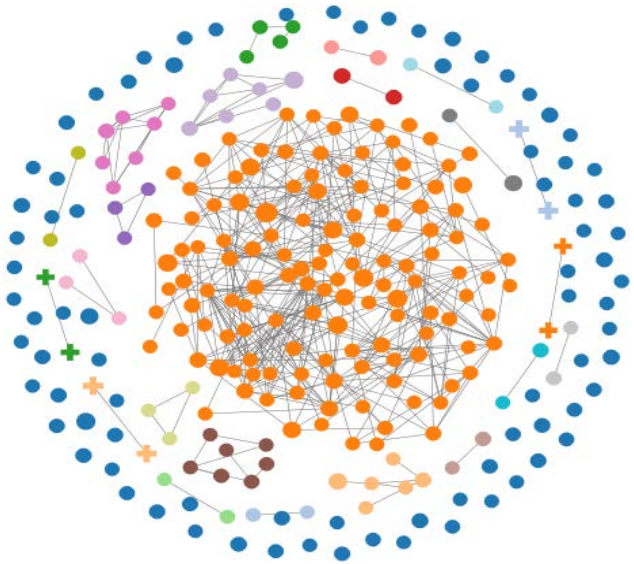
It’s Autonomous

- GHOST unburdens CDC staff
- Actionable intelligence is provided automatically to users
- CDC analysts can focus on the population picture rather than individual investigations

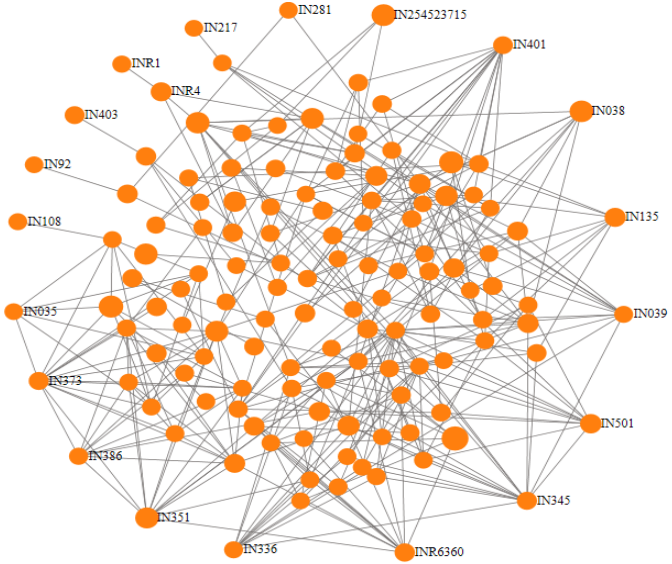
Traditional Phylogenetic Analysis



GHOST Analysis



Major Cluster



GHOST was an essential part of the "The Indiana Injection Drug Use-Associated HIV/HCV Outbreak Response Team", which was selected as the winner of the 2015 CDC/ATSDR Honor Award for Excellence in Emergency Response – Domestic.

It’s Modular

- We are currently testing new HCV modules that
  - Detect Persons Who Inject Drugs
  - Estimate duration of infection
  - Monitor for drug resistance
- Modules to investigate Hepatitis A and B are in the works
- Discussions with Malaria and HIV Divisions to integrate these other pathogens are in progress

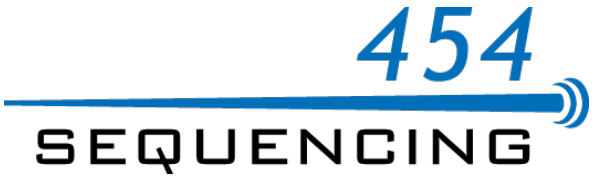
It bridges worlds

- Regional sequencing laboratories are in the works; called GHOST centers
- Pooling of samples drives down laboratory costs
- Laboratory robots will further reduce costs and errors

These are the shadows of things that have been.

GHOST of Infrastructures Past

454 Jr. Pipeline



- Produces approximately 3 Mb files
- Correction algorithm concerned primarily with fixing homo-polymer errors
- Implemented in a mixture of MATLAB and Java

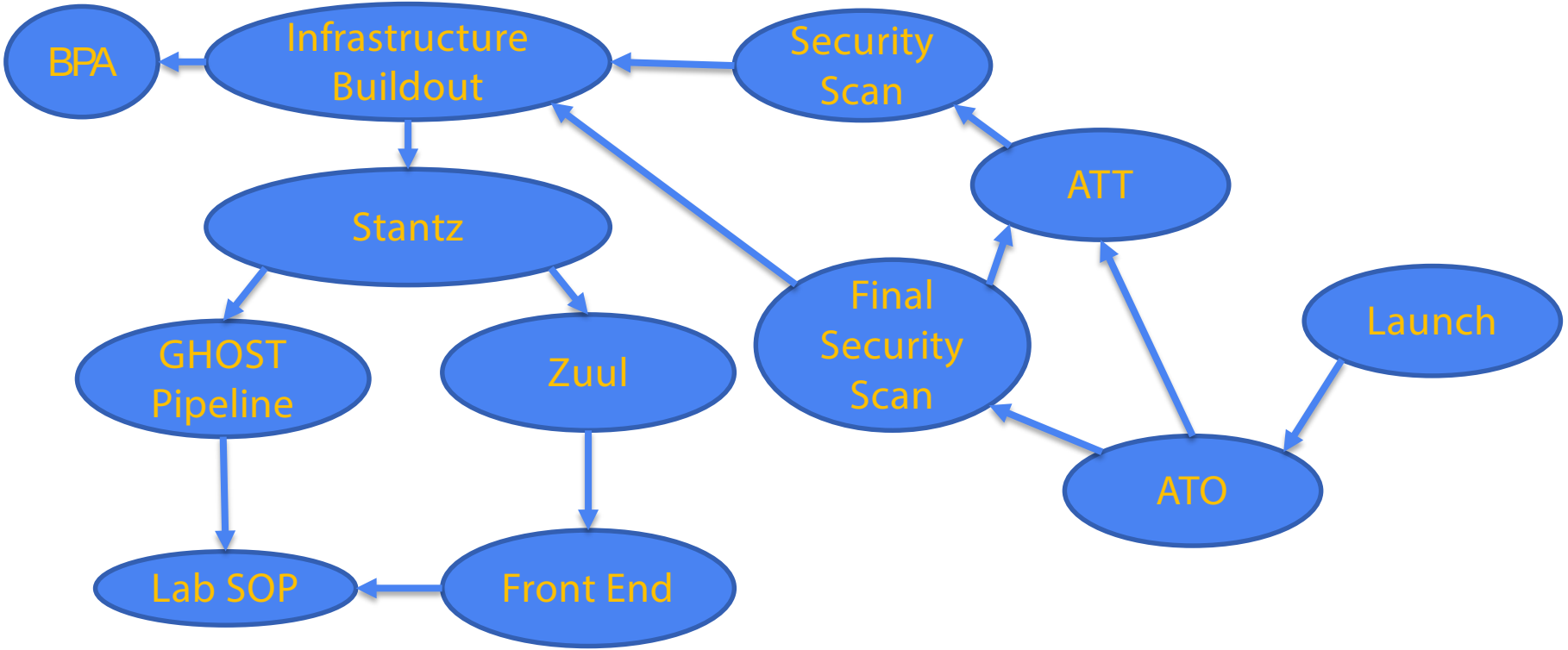
Infrastructure

- Runs on Velocity - a Windows-based server acquired by ITSO R&D
- Housed in Bldg. 21 datacenter
- 96 Gb Ram, 4x 2.27 GHz Intel Xeon Processors

Come in and know me better, man

GHOST of Infrastructures Present

Project Dependency Graph



Backend software development

- A “Command and Control” server named Stantz is being developed
- Stantz waits for incoming tasks and launches appropriate work to complete the task.
- Stantz also “keeps notes” about what should be going on and attempts to recover automatically from faults.

- Zuul, a bridge between the front-end and back-end, was developed by the NCHHSTP informatics office.
- Zuul handles transfer of data between the servers that host the website and the computational cluster.

MiSeq Pipeline

- Sequencer files approximately 1.4 Gb
- Each file contains approximately 1.1 million reads
- Correction algorithm concerned primarily with correcting substitution errors
- Implemented in a mixture of Python, Cython, and command line programs (mafft, hmmer, blast+)

Cloud infrastructure

- 6 Elastic Compute Cloud nodes
  - 2x t2.large head node and master node
  - 24x t2 medium
- 2 Simple Queueing Service queues
- 1 Simple Storage Service object bucket
- Runs Grid Engine using DRMAA to interface

Shadows of things that will be

GHOST of Infrastructure Yet to Come

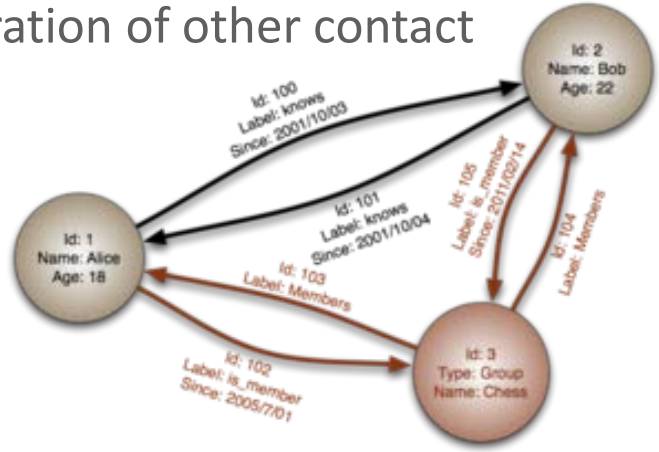
Sample Storage and Survillence

- We are looking into developing a backend data manager service named Gozer
- This would deal with sample databases, meta-data, format conversions, permissions, access auditing, etc.

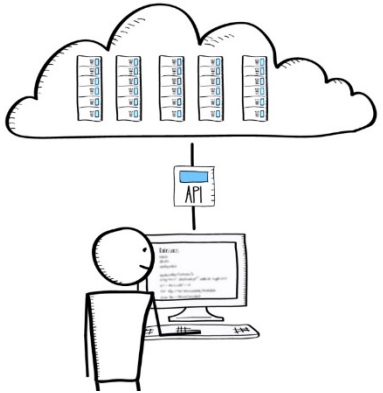
- Vinz will be a middle-ware similar to Zuul that is concerned with surveillance driven tasks.
- Vinz will be on the lookout for anomalies in the database, such as transmission events between different studies

Link Database (OrientDB)

- One way to improve the efficiency and responsiveness of GHOST is to integrate a technology called “Graph Databases”
- When a new network tracing task comes in we would query the database for any links we have already measured, and only trace new ones.
- The graph database would also allow for integration of other contact tracing information into the analysis.



Scalability and System Simulation

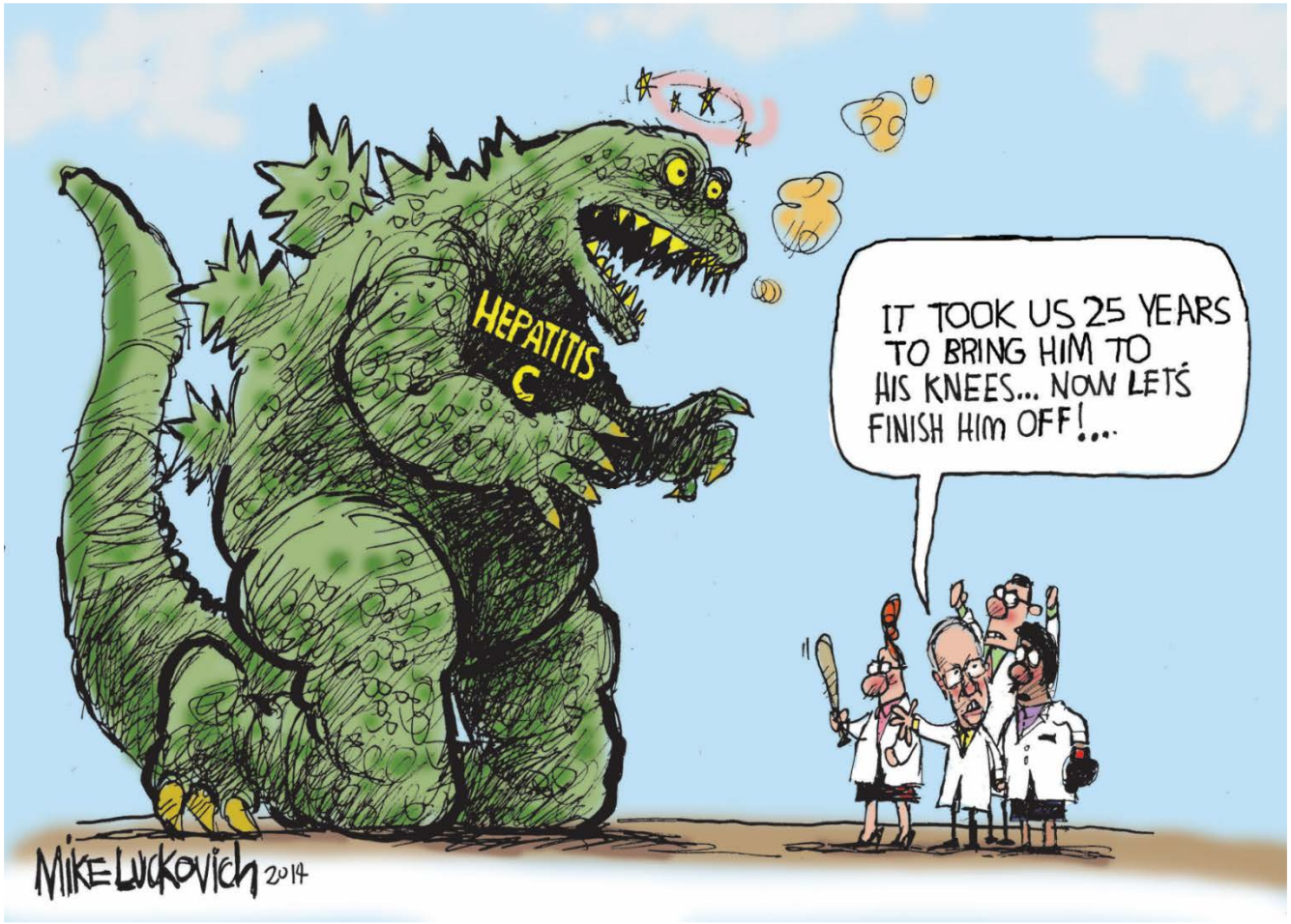


- The current implementation uses a fixed number of compute nodes
- cfnCluster will automatically create new nodes to meet demand.
- Currently working with consultant for this effort

- Tobin spirit guide is a discrete-time queueing simulator of the GHOST system
- It will allow us to simulate the system to measure things such as turn around time, system cost, maximum number of users, etc.

Thanks

- DVH – Yury Khudyakov, David Campo, Zoya Dimitrova, Pavel Skums, Amanda Sue, Garrett Longmire, Lilia Ganova-Raeva, Maggie Medrzycki, Sumathi Ramachadran, Natalia Saveleva, Hong Thai, Gilberto Vaughan, Guo-liang Xia, Saleem Kamili
- DVH Alumni – Inna Rytsareva, Hajung Roh
- NCHHSTP Informatics Office – Thom Sukalac, Robin Tracy, Max Mirabito, Silver Wang
- ITSO Application Hosting Branch – Chad McCown, Victor Bolet, Van Vongsaphanh, Jim Yohn, Stephen Long, Jamie Saucier, Rob Bahurinsky, Steve Zach, Barbara Boone
- NCHHSTP IT Security Team – Bill Dolan, Ralph Vaughn


HEPATITIS C
IT TOOK US 25 YEARS TO BRING HIM TO HIS KNEES... NOW LETS FINISH HIM OFF!...
MIKE LUCKOVICH 2014

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.