SPHINX: A system for telling computers and humans apart through audio CAPTCHA

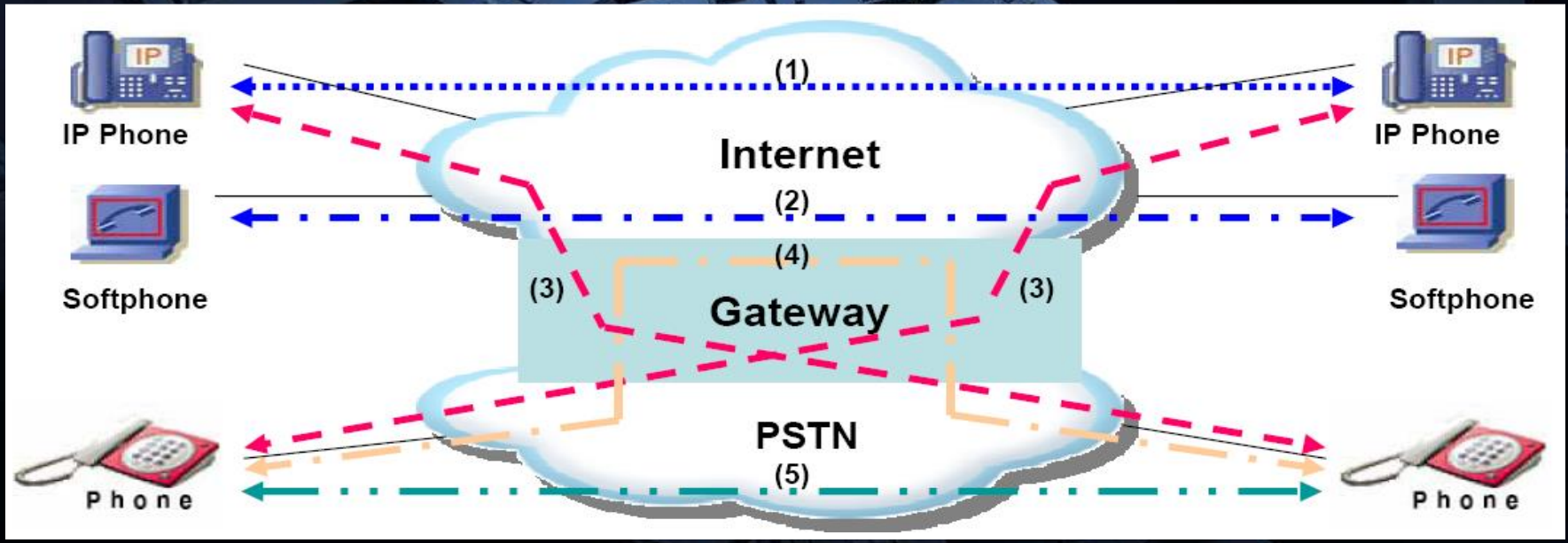
Yannis Soupionis

Outline

- Introduction
  - Internet Telephony
  - Spam over Internet Telephony (SPIT)
  - SPIT Phenomenon
- Methodology
- Research approach
  - Security Policies
  - CAPTCHA
  - Formal Verification
- Contribution – Future research

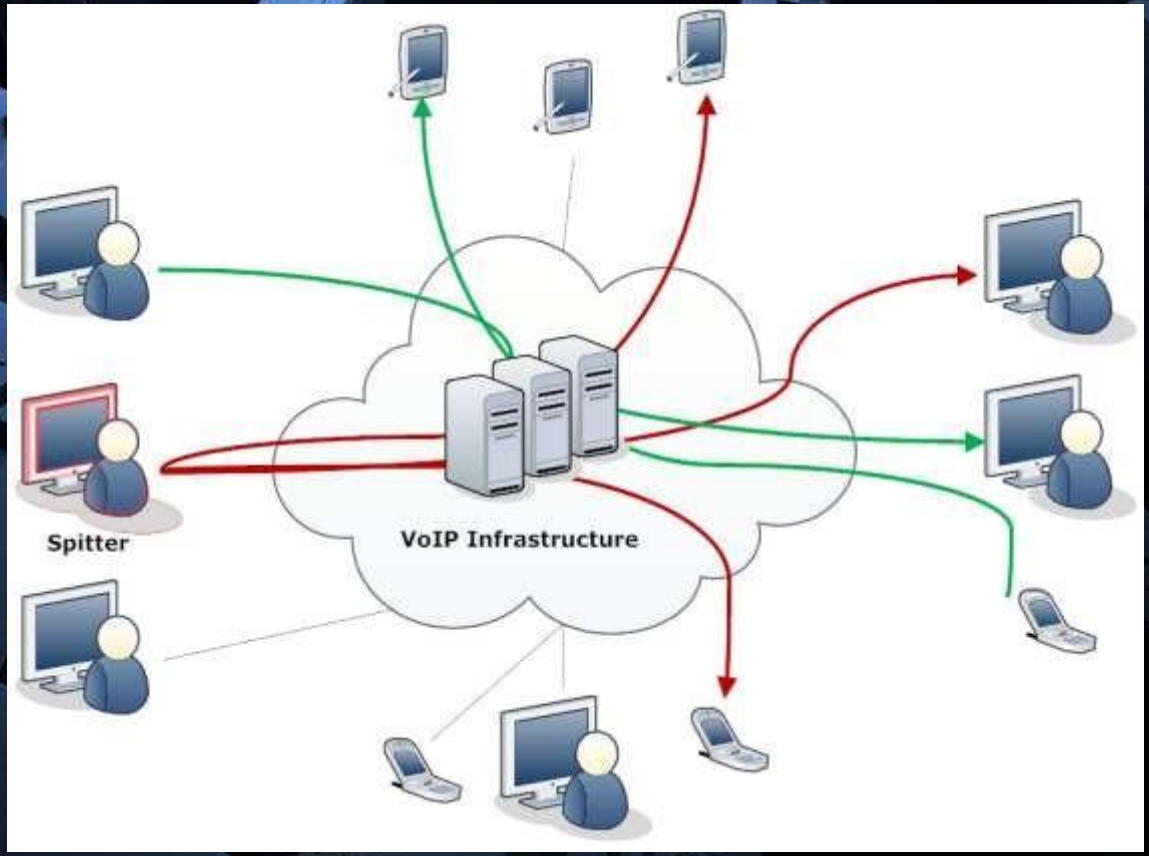
Voice-over-IP (VoIP)

- **Data networks** and **voice networks** convergence
- **Voice-over-IP (VoIP) technologies** stand as Internet telephony infrastructure.
- Based on protocols, such as the **Session Initiation Protocol (SIP)** for signaling phase and the **RTP** for transmiting voice or multimedia content.



SPam over Internet Telephony (SPIT)

- Bulk unsolicited set of sessions
  - Call initiations
  - Instant messages
  - Presence requests



Rosenberg J., Jennings C., The Session Initiation Protocol (SIP) and Spam, Technical Report RFC 5039, Network Working Group, January 2008

SPIT phenomenon

- Implementation of mechanisms for tackling SPIT attacks by well-known companies as NEC and Microsoft.
- Recorded SPIT attacks
  - 4 million spam texts sent every day - telegraph.co.uk
  - Stop Spam And Unwanted Calls - cbsnews.com
- Environmental burden due to SPAM/SPIT
  - Carbon Footprint of Spam ≈ 3 million cars - thegreenitreview.com - McAfee
- Economic benefits in response rates to SPAM ≈ 0,00001% - ACM CCS 2008

Email vs. Voice Spam

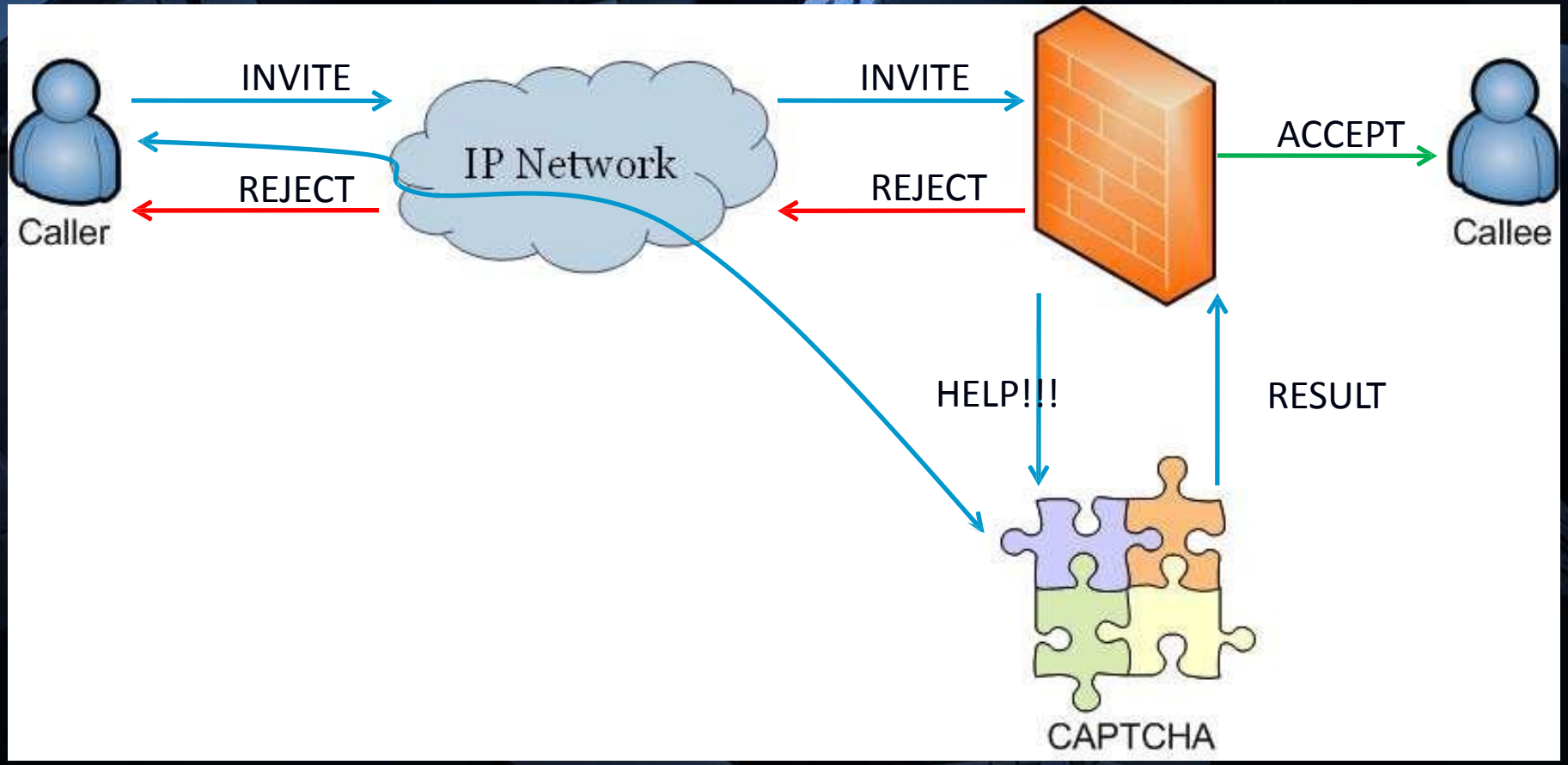
Similarities

- **Common incentives,** e.g. seeking financial gain or influence.
- **Common** implementation techniques, e.g. automatic production of mass low cost messages/calls, use of real end-users' addresses, address collection etc.

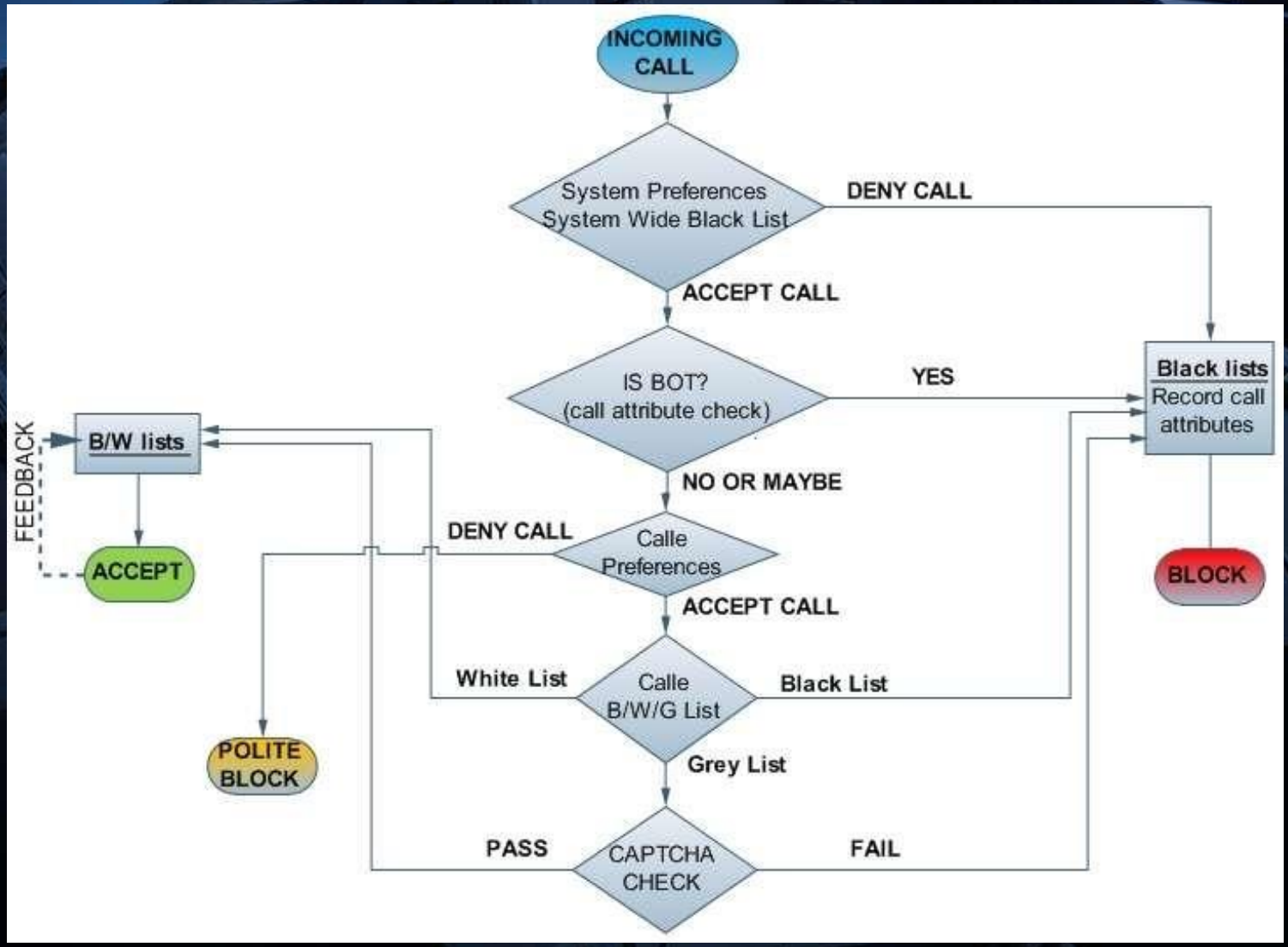
Differences

- Communication by email is essentially **asynchronous**, while VoIP communication is mainly **synchronous**.
- In the VoIP environment unreasonable delays **are not technically acceptable.**
- Spam email is mainly composed of **text** (perhaps images as well) while SPIT is primarily composed by **sound** and **image** (far less by text).
- A SPIT call usually creates more intensive **disturbance** to the user.

Methodology

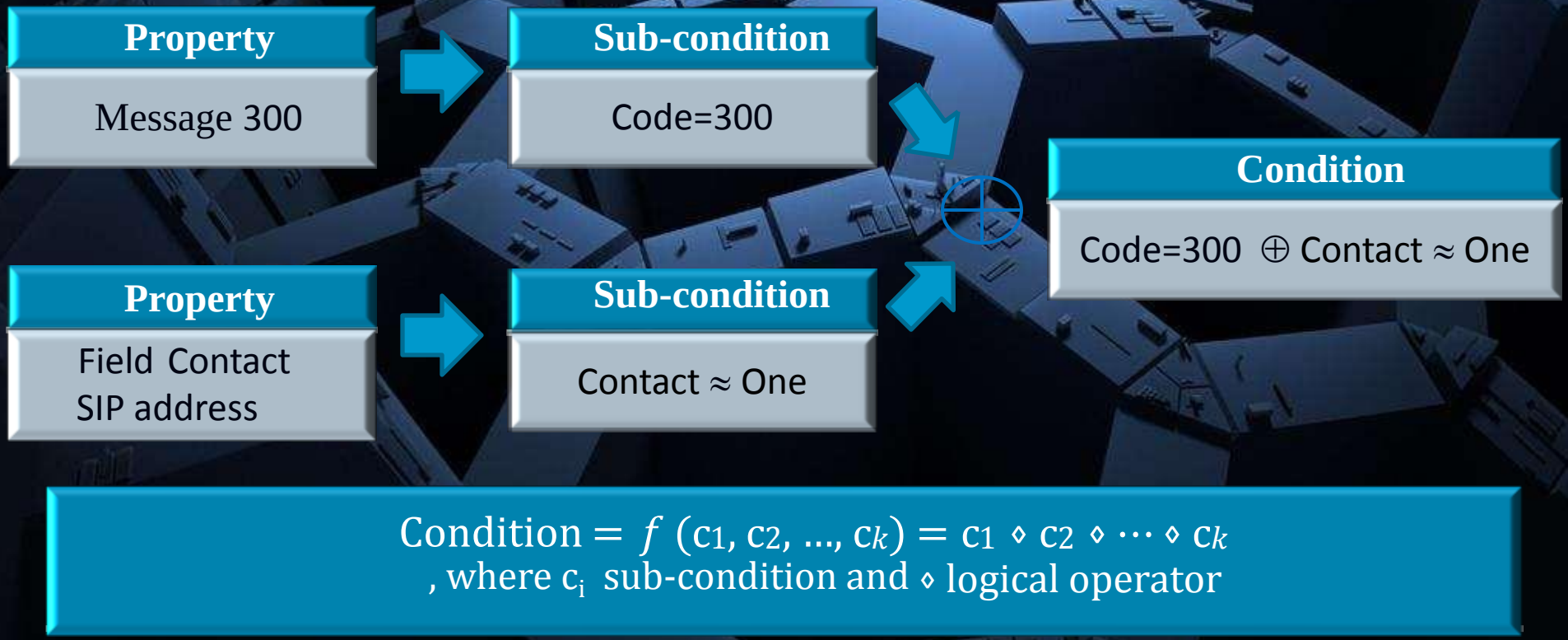


Security Policy



Security Policy Implementation

Scenario	The device (UAC) receives an answer with a code 300 (Multiple Choice), while in the Contact field of the answer indicates a SIP address



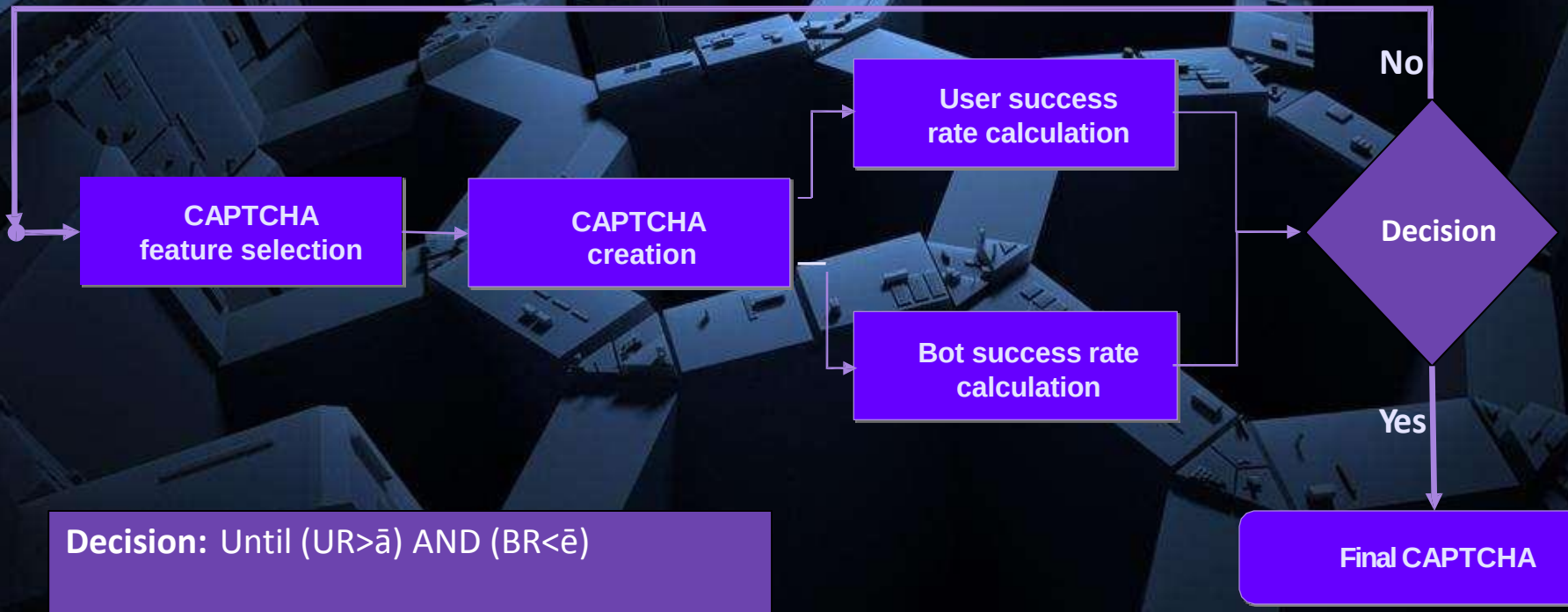
$$\text{Condition} = f(c_1, c_2, \ldots, c_k) = c_1 \diamond c_2 \diamond \cdots \diamond c_k$$

, where c_i sub-condition and $\diamond$ logical operator

Soupionis Y., Gritzalis D., "ASPF: An adaptive anti-SPIT policy-based framework", in Proc. of the 6th International Conference on Availability, Reliability and Security (ARES-2011), Pernul G. (Ed.), pp. 153-160, Austria, August 2011.

CAPTCHA

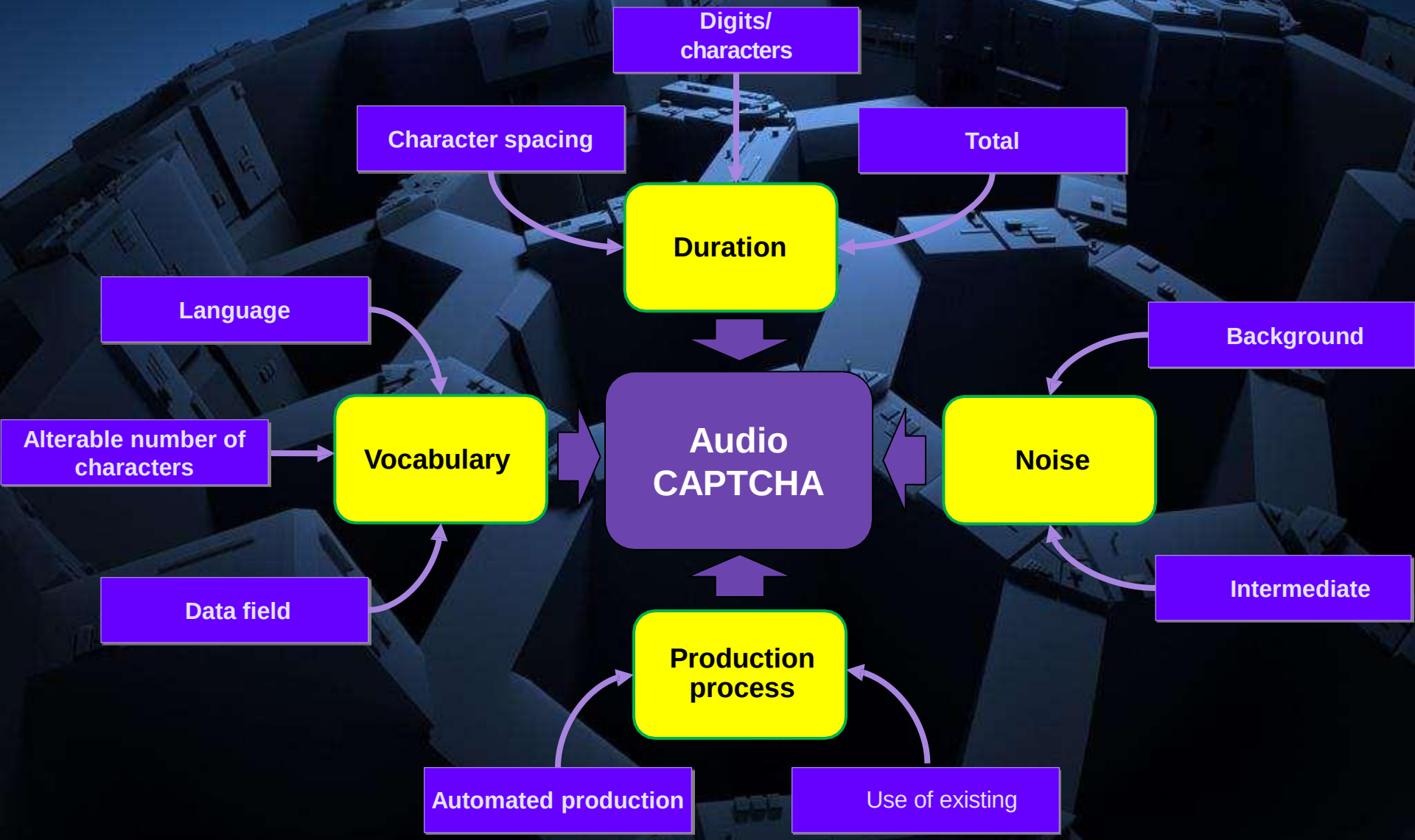
- Completely Automated Public Tests to tell Computers and Humans Apart



Decision: Until (UR>ā) AND (BR<ē)

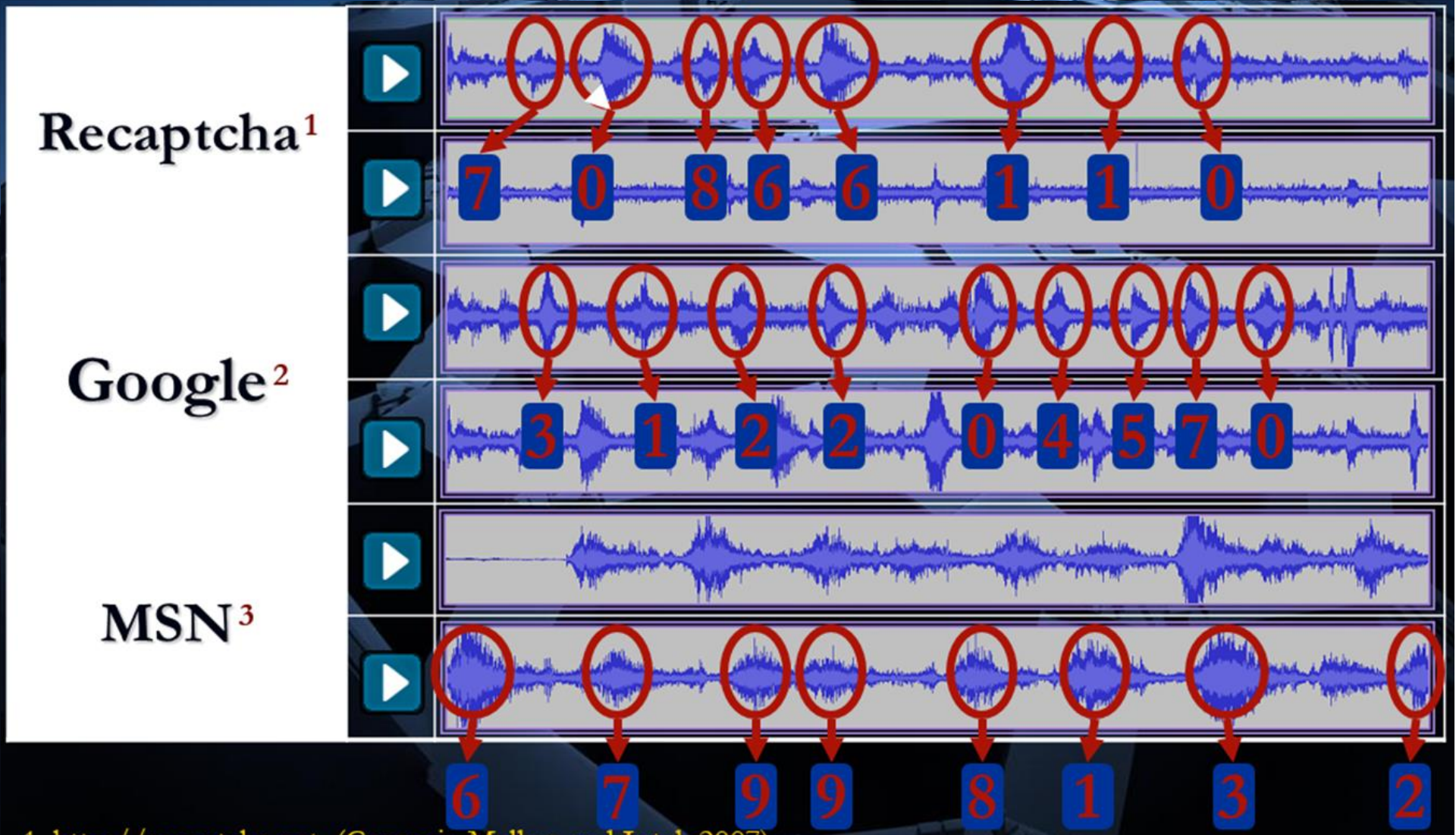
UR: User success rate
BR: Bot success rate
ā: User rate threshold
ē: Bot rate threshold

CAPTCHA characteristics



Soupionis Y., Tountas G., Gritzalis D., "Audio CAPTCHA for SIP-based VoIP", Proc. of the 24th International Information Security Conference, pp. 25-38, Springer (IFIP AICT 297), Cyprus, 2009.

Current Audio CAPTCHA



1. http://recaptcha.net (Carnegie Mellon and Intel, 2007)
2. http://gmail.com (Google, 2008) **(Vorm bot access rate: 33%)**
3. https://accountservices.passport.net/reg.srf (Microsoft, 2008) **(Vorm bot access rate: 75%)**

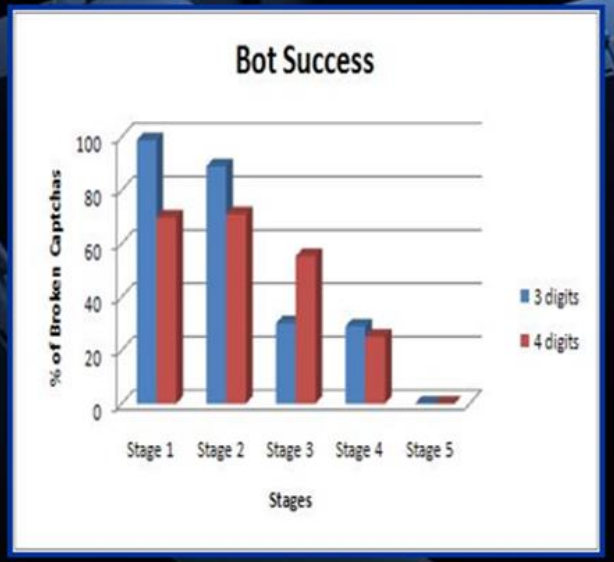
Comparison of existing solutions of audio CAPTCHA

Audio CAPTCHA / Characteristics	Google	MSN	Recaptcha	eBay	Secure image captcha	Mp3Captcha	Captchas. net	bokelman	slashdot	Authorize	AOL	Digg
User's Success rate	60%	80%	50%	95%	98%	98%	98%	98%	95%	95%	95%	95%
Background noise	Voice, sound	Voice, sound	Sound	Voice, sound	Sound	No	No	No	No	No	Voice	Sound
Intermediate noise	Sound	Sound	No	No	No	No	No	No	No	No	Sound	No
Data field	0-9	0-9	Words	0-9	A-Z, a-z, 0-9	A-Z, a-z, 0-9	a-z, 0-9	A-Z, a-z, 0-9	Words	A-Z, a-z, 0-9	A-Z, a-z, 0-9	A-Z, a-z, 0-9
Number of characters in a snapshot	5-10	10	10-20	6	4	4	6	4	<9	5	8	5
Rare reappearance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
Production process	Unknown	Unknown	Unknown	Unknown	Automated	Automated	Automated	Automated	Unknown	Unknown	Unknown	Unknown
Speaker voice	Multiple languages	Multiple languages	en	Multiple languages	en	en, fr, it, de	en, de, it, nl, fr	en	en	en	en	en
Different speakers	Yes	No	Yes	No	Yes	No	No	No	No	No	Yes	No
Duration(sec)	0:10-0:15	0:05-0:09	~0:04	~0:04	~0:04	~0:04	~0:08	0:04-0:05	0:03-0:04	0:05	0:10	0:08

Soupionis Y., "SPAM prevention in VoIP networks via security policies and audio CAPTCHA"., PhD Thesis, Dept. Of Informatics, Athens University of economics and Business, Greece, 2011.

Audio CAPTCHA implementation

	Number of speakers	Time delay	Intermediate noise	Background noise	Number of training snapshots
Phase 1	1	x	x	x	20
Phase 2	3	x	x	x	50
Phase 3	5	x	x	☑	100
Phase 4	7	☑	x	☑	100
Phase 5	7	☑	☑	☑	100



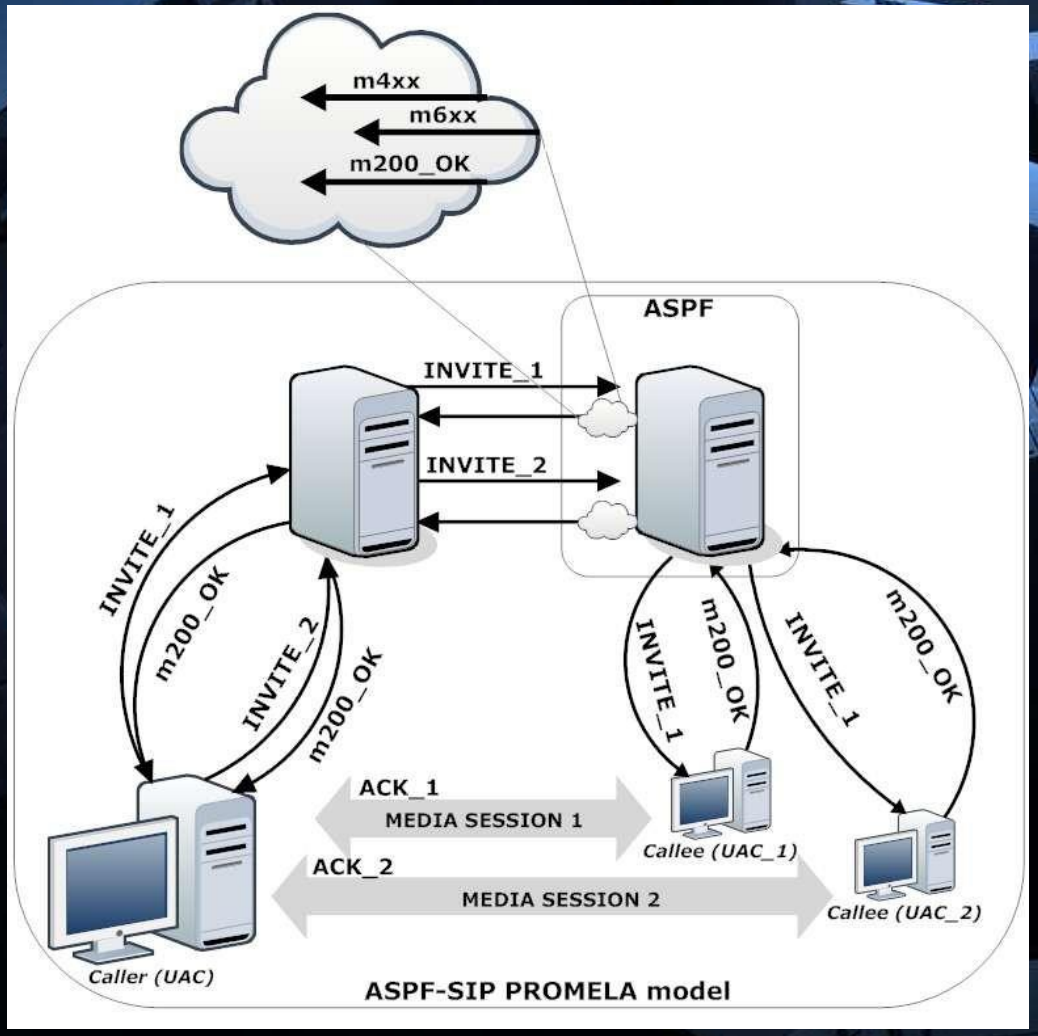
Soupionis Y., Gritzalis D., "Audio CAPTCHA: Existing solutions assessment and a new implementation for VoIP telephony", Computers & Security, Vol. 29, No. 5, pp. 603-618, 2010.

Formal verification

- Software system analysis via **mathematical techniques**, where the examined system is represented in an abstract level, can **verify diverse groups of properties.**
  - Intel Pentium (1994) -> an error (bug) at the floating point hardware of the microprocessor-> Cost 400 million $
- Formal verification of communication protocols
  - Testing and analysis of the entire state graph which is produced by a model/system.
- Correctness properties
  - Assertions or temporal logic formulae that are algorithmically validated by state exploration across all possible execution paths
- Modeling tool
  - SPIN – popular and open source

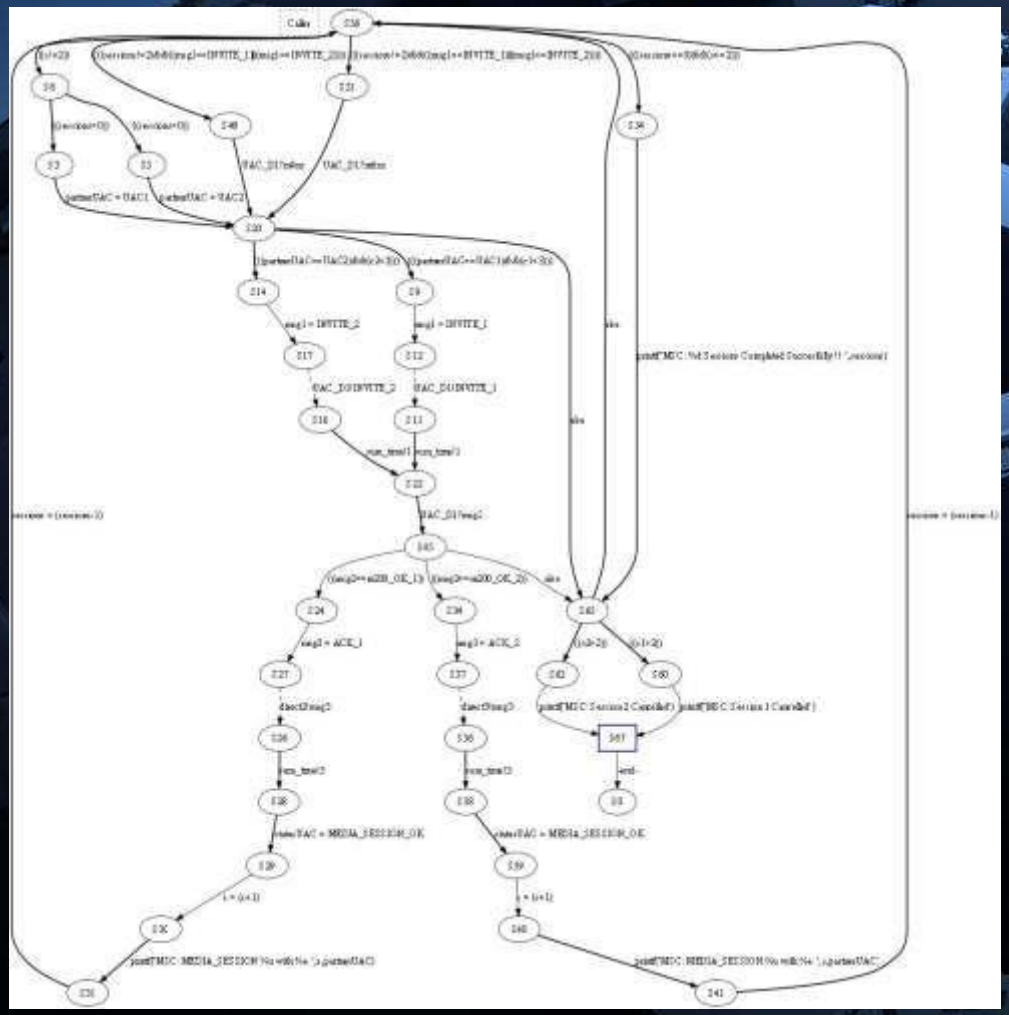
E. Clarke, A. Emerson, and J. Sifakis, "Model Checking: Algorithmic Verification and Debugging", *ACM 2007 Turing Award*
INTEL FDIV Replacement Program ,California., December, 1994 (http://www.intel.com/support/processors/pentium/fdiv/)

SPIN model



- Message codes
  - INVITE
  - 2xx successful answer
  - 3xx redirecting response
  - 4xx request failure
  - 6xx system failure
  - ACK
- Message codes
  - 1xx optionally

SPIN model



Property

Correlation operator

Q1: [](q→p)

Time operator

time<4000

Q1: [](q→p)

sessions==0

Verification results

- Absence of deadlock
  - Executions either terminate with successfully completed initiated sessions or with failed sessions, due to dispatched messages that declare an error.
- Call establishment timeliness for all error-absent execution paths

Property description	States	Transision	Memory (MB)
Session establishment < 4000 ms	3.8e+06	7.181e+06	585.309
Parallel session establishment < 6500 ms	3.8e+06	7.246e+06	616.11
Full state graph (Absence of deadlock)	3.8e+06	7.181e+06	585.309

Soupionis Y., Basagiannis S., Katsaros P., Gritzalis D., "A formally verified mechanism for countering SPIT", in Proc. of the 5th International Conference on Critical Information Infrastructure Security (CRITIS-2010), Springer, Greece, September 2010.

Gritzalis D., Katsaros P., Basagiannis S., Soupionis Y., "Formal analysis for robust anti-SPIT protection using model-checking", International Journal of Information Security, Vol. 11, No. 2, pp. 121-135, 2012.

Conclusions

- ✓ The VoIP widespread use introduces not only many **benefits**, but also **new threats**.
- ✓ The adequate mitigation of SPIT requires **multi-factorial approach** (Policies & CAPTCHA) – The existing anti-spam techniques are **not sufficient**.
- ✓ The anti-SPIT techniques should aim to discover, identify and tackle more and **new kinds of attacks**.
- ✓ The audio CAPTCHA that capitalizes the voice **tone**, **intermediate noise** and the digits and noise **randomly distribution** within the message is encouraging resistance against bots.

References

1. Dritsas S., Tsoumas B., Dritsou V., Konstantopoulos, P., Gritzalis D., "OntoSPIT: SPIT Management through Ontologies", *Computer Communications*, Vol. 32, No. 2, pp. 203-212, 2009.
2. Gritzalis D., Katsaros P., Basagiannis S., Soupionis Y., "Formal analysis for robust anti-SPIT protection using model-checking", *International Journal of Information Security*, Vol. 11, No. 2, pp. 121-135, 2012.
3. Soupionis Y., Basagiannis S., Katsaros P., Gritzalis D., "A formally verified mechanism for countering SPIT", in Proc. of the 5th International Conference on Critical Information Infrastructure Security (CRITIS-2010), Wolthusen S., et al. (Eds.), pp. 128-139, LNCS-6712, Springer, Greece, September 2010.
4. Gritzalis D., Mallios J., "A SIP-based SPIT management framework", *Computers & Security*, Vol. 27, No. 5-6, pp. 136-153, 2008.
5. Gritzalis D., Marias G., Rebahi Y., Soupionis Y., Ehlert, S., "SPIDER: A platform for managing SIP-based spam over Interent Telephony", *Journal of Computer Security*, Vol. 19, No. 5, pp. 835-867, 2011.
6. Kandias M., Virvilis N., Gritzalis D., "The insider threat in Cloud Computing", *Proc. of the 6th International Workshop on Critical Infrastructure Security*, pp. 93-103, Springer (LNCS 6983), Switzerland, 2011.
7. Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., "An Insider Threat Prediction Model", *Proc. of the 7th International Conference on Trust, Privacy and Security in Digital Business*, pp. 26-37, Springer (LNCS 6264), Spain, 2010.
8. Soupionis Y., Gritzalis D., "ASPF: An adaptive anti-SPIT policy-based framework", *Proc. of the 6th International Conference on Availability, Reliability and Security*, pp. 153-160, Austria, 2011.
9. Soupionis Y., Tountas G., Gritzalis D., "Audio CAPTCHA for SIP-based VoIP", *Proc. of the 24th International Information Security Conference*, pp. 25-38, Springer (IFIP AICT 297), Cyprus, 2009.
10. Soupionis Y., Dritsas S., Gritzalis D., "An adaptive policy-based approach to SPIT management", *Proc. of the 13th European Symposium on Research in Computer Security*, pp. 446-460, Springer, Spain, 2008.
11. Soupionis Y., Gritzalis D., "Audio CAPTCHA: Existing solutions assessment and a new implementation for VoIP telephony", *Computers & Security*, Vol. 29, No. 5, pp. 603-618, 2010.
12. Stachtiari E., Soupionis Y., Katsaros P., Mentis A., Gritzalis, D., "Probabilistic model checking of CAPTCHA admission control for DoS resistant anti-SPIT protection", *Proc. of the 7th International Conference on Critical Information Infrastructures Security*, Springer (LNCS 7722), Norway, 2012.
13. Tassidou A., Efraimidis P., Soupionis Y., Mitrou L., Katos V., "User-centric privacy-preserving adaptation for VoIP CAPTCHA challenges", *Proc. of the 6th International Symposium on Human Aspects of Information Security and Assurance*, Greece, 2012.